

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 13:03:17 ; Search time 3909 Seconds
 (without alignments) 10381.773 Million cell updates/sec

Title: US-09-990-726-222

Perfect score: 992

Sequence: 1 ggacagccaggaaactagg.....aaaaaaaaaaaaaaaaaa 992

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapet 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Maximum Match 0%
 Listing First 150 summaries

Database : GenEmbl:
 1: gb_ba: *
 2: gb_htg: *
 3: gb_in: *
 4: gb_com: *
 5: gb_ov: *
 6: gb_pat: *
 7: gb_ph: *
 8: gb_pi: *
 9: gb_pr: *
 10: gb_to: *
 11: gb_sts: *
 12: gb_sy: *
 13: gb_un: *
 14: gb_vl: *
 15: em_ba: *
 16: em_fun: *
 17: em_hum: *
 18: em_ir: *
 19: em_mu: *
 20: em_om: *
 21: em_or: *
 22: em_un: *
 23: em_vl: *
 24: em_ph: *
 25: em_p1: *
 26: em_co: *
 27: em_sts: *
 28: em_ov: *
 29: em_vl: *
 30: em_htg_hum: *
 31: em_htg_inv: *
 32: em_htg_other: *
 33: em_htg_mus: *
 34: em_htg_p1n: *
 35: em_htg_r0d: *
 36: em_htg_man: *
 37: em_htg_vrl: *
 38: em_sy: *
 39: em_htg_hum: *
 40: em_htg_mus: *
 41: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a

OM	nucleic	nucleic	Result				Description			
			No.	Score	Query %	Match	Length	DB	ID	
			1	992	100.0	992	6	AX252529	Sequence	
			2	992	100.0	992	6	AX358896	Sequence	
			3	992	100.0	992	6	AX362389	Sequence	
			4	992	100.0	992	6	AX403335	Sequence	
			5	504.6	50.9	505	6	BD107970	EST and e	
			6	501.6	50.6	502	6	AX588820	Sequence	
			7	403.8	41.3	414	11	G60502	SHG-33533	
			8	300.4	30.3	154440	2	AC032035	Homo sapi	
			9	300.4	30.3	177738	9	AC021593	Homo sapi	
			10	300.4	30.3	209751	9	AC087645	Homo sapi	
			11	291.8	29.5	229426	2	AC010532	Homo sapi	
			12	291.2	29.4	154940	2	AC032035	Homo sapi	
			13	179.4	18.1	229426	2	AC10532	Homo sapi	
			14	163.4	16.5	116300	10	AL645856	Mouse DNA	
			15	163.4	16.5	197194	2	AL954690	Mus muscu	
			16	160.2	16.1	49532	2	AC100565	Mus muscu	
			17	149	15.0	150	11	G43545	WIF-2405-S	
			18	149	15.0	150	11	G43546	WIF-2407-S	
			19	127.8	12.9	161043	2	AC143361	Macaca mu	
			20	69.2	7.0	151750	9	AC003666	Homo sapi	
			21	69.2	7.0	208670	9	AC103388	Homo sapi	
			22	62.8	6.3	177921	9	AL537575	Human DNA	
			23	61.8	6.2	146883	9	AC018788	Homo sapi	
			24	61	6.1	120200	9	AF429315	Homo sapi	
			25	60.8	6.1	164762	2	AC016177	Homo sapi	
			26	60.8	6.1	170898	9	AL353764	Human DNA	
			27	60.4	6.1	129770	9	AC104684	Homo sapi	
			28	60.4	6.1	171655	2	AC067919	Homo sapi	
			29	58.8	5.9	123437	9	AC010072	Homo sapi	
			30	58.6	5.9	151993	9	AC093574	Homo sapi	
			31	58.6	5.9	19635	9	AC012087	Homo sapi	
			32	58.4	5.9	4950	9	HSIM130AC2		
			33	58.4	5.9	177260	9	AC131206	Homo sapi	
			34	58	5.8	140055	2	AC023003	Homo sapi	
			35	58	5.8	151553	9	AL139342	Human DNA	
			36	58	5.8	163085	9	AC019159	Homo sapi	
			37	57.8	5.8	99268	9	AC091158	Homo sapi	
			38	57.8	5.8	119853	9	AL136314	Human DNA	
			39	57.8	5.8	138370	9	AL354713	Human DNA	
			40	57.8	5.8	156857	2	AC025375	Homo sapi	
			41	57.8	5.8	161940	2	AC079289	Homo sapi	
			42	57.4	5.7	173270	9	AC025522	Homo sapi	
			43	57.4	5.7	164405	9	AC107054	Homo sapi	
			44	57.4	5.7	166658	2	AC025258	Homo sapi	
			45	57.4	5.7	183778	2	AC02104	Homo sapi	
			46	56.6	5.7	175947	2	AC145171	Homo sapi	
			47	56.6	5.7	151841	9	AC013408	Homo sapi	
			48	56.4	5.7	254455	2	AC099922	Homo sapi	
			49	55.8	5.6	112389	2	AC073598	Homo sapi	
			50	56	5.6	189271	9	AL138752	Human DNA	
			51	56	5.6	194347	2	AC091575	Homo sapi	
			52	56	5.6	207069	9	AC090922	Homo sapi	
			53	56	5.6	254455	2	AC099902	Homo sapi	
			54	55.8	5.6	193078	9	AC099780	Homo sapi	
			55	55.8	5.6	193793	9	AP006232	Homo sapi	
			56	55.4	5.6	214666	9	AC069544	Homo sapi	
			57	54.6	5.5	169856	2	AL353559	Homo sapi	
			58	54.6	5.5	181356	2	AC069441	Homo sapi	
			59	54.6	5.5	183104	9	AL355601	Human DNA	
			60	54.6	5.5	190978	9	AC099780	Homo sapi	
			61	54.6	5.5	208026	9	AP006232	Homo sapi	
			62	54.4	5.5	207190	9	AF295017	Homo sapi	
			63	54.4	5.5	217898	9	AC008498	Homo sapi	
			64	54.2	5.5	106340	9	AC097104	Homo sapi	
			65	54.2	5.5	146017	2	AC027473	Homo sapi	

AC02035 154840 bp DNA linear HTG 08-SEP-2000
Homo sapiens chromosome 17 clone RP1-141D15 map 17, *** SEQUENCING
DEFINITION

AC02035 18 unordered pieces.
AC02035.3 GT:9994161
VERSION
KEYWORD HTG; PHAS1.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

Birren, B.; Linton, L.; Nusbaum, C.; Lander, E.; Abraham, H.; Allen, N.;
Anderson, S.; Baldwin, J.; Barna, N.; Bastien, V.; Beda, F.;
Boguski, M.; Brown, A.; Burkett, G.;
Campopiano, A.; Castle, P.; Chopek, Y.; Colangelo, M.; Collins, S.;
Collimore, A.; Cooke, P.; DeAngelis, K.; Dewart, K.; Diaz, J.S.;
Dodge, S.; Domino, M.; Doyle, M.; Ferreira, P.; Fitzgerald, W.; Gage, D.;
Galagan, J.; Gardyna, S.; Ginde, S.; Goyette, M.; Graham, L.;
Grand-Pierre, N.; Grant, G.; Hagos, B.; Heaford, A.; Horton, L.;
Howland, J.C.; Iliev, I.; Jones, R.; Johnson, R.; Jones, C.; Kahn, L.; Karatas, A.;
Klein, J.; LaRocque, K.; Lamazares, R.; Landers, T.; Lehozcy, J.;
Levine, R.; Lieu, C.; Liu, G.; Locke, K.; Macdonald, P.; Marquis, N.;
McCarthy, M.; McEwan, P.; McGuire, K.; McPherrers, R.;
Meldrim, J.; Menuez, L.; Mihova, T.; Miranda, C.; Mlenga, V.; Morrow, J.;
Murphy, T.; Naylor, J.; Norman, C.H.; O'Conor, T.; O'Donnell, P.;
O'Neill, D.; Oliver, T.M.; Oliver, J.; Peterson, K.; Pierre, N.;
Pisan, C.; Pollara, V.; Raymond, C.; Riley, R.; Rogov, P.; Rothman, D.;
Roy, A.; Santos, R.; Schauer, S.; Severy, P.; Spencer, B.;
Strange-Monahan, N.; Stojanovic, N.; Subramanian, A.; Talman, J.J.;
Tessyay, S.; Theodore, J.; Tirrell, A.; Travers, M.; Trigilio, J.;
Vassiliev, H.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Ye, W.J.;
Young, G.; Zainoun, J.; Zimmer, A. and Zody, M.

DIRECT SUBMISSION

Submitted (03-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02142, USA
On Sep 8, 2000 this sequence version replaced gi:7705196.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project information: <http://www-seq.wi.mit.edu/>

Center Project name: L9138
Center Clone name: 141_D_15

NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2373: contig of 2373 bp in length
* 2374 2473: gap of 100 bp
* 2474 4930: contig of 2457 bp in length
* 4931 5030: gap of 100 bp
* 5031 7215: contig of 2185 bp in length
* 7216 7315: gap of 100 bp
* 7316 9541: contig of 2226 bp in length
* 9542 9641: gap of 100 bp
* 9642 12211: contig of 2570 bp in length
* 12212 12311: gap of 100 bp
* 12312 14923: contig of 2612 bp in length
* 14924 15023: gap of 100 bp
* 15024 19094: contig of 4071 bp in length

1 19095 19196: gap of 100 bp
* 19195 22288: contig of 3092 bp in length
* 22287 22387: gap of 100 bp
* 22387 26147: contig of 3761 bp in length
* 26148 26248: gap of 100 bp
* 26248 31933: contig of 5686 bp in length
* 31934 32033: gap of 100 bp
* 32033 37334: contig of 5306 bp in length
* 37334 37439: gap of 100 bp
* 37440 42858: contig of 5419 bp in length
* 42859 42959: gap of 100 bp
* 42959 48810: contig of 5832 bp in length
* 48811 48910: gap of 100 bp
* 48911 56020: contig of 7119 bp in length
* 56020 56120: gap of 100 bp
* 56120 62189: contig of 6060 bp in length
* 62120 62289: gap of 100 bp
* 62289 82957: contig of 20870 bp in length
* 82957 83056: gap of 100 bp
* 83056 107059: contig of 24000 bp in length
* 107060 107160: gap of 100 bp
* 107160 154845: contig of 47681 bp in length.

FEATURES

source 1..154845
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP1-141D15"
/clone_id="RPCI-11_Human_Male_BAC"
/map="17"
/clone_id="41741_C_41557_g_32874_t_2488_others"

BASE COUNT

ORIGIN

Query Match 36180 a 36180 a 30.3%; Score 300 4; DB 2; Length 154840;
Best Local Similarity 99.7%; Pred. No. 1e-61; Matches 301; Mismatches 1; Indels 0; Gaps 0;

Qy 131 AGATAATACCCCTGTGTCATTGCTCAAGTGCTGAGTTTCCTCCAAAGGCCG 190
Db 1901 AGAAATTAACCCCTGTGTCATTGCTCAAGTGCTGAAAGTTTCCTCCAAAGGCCG 1842
Qy 191 CTGGGTGTCATAAACCGTCTGAGCACCACGCCGCACTACACTATTCCTCTG 250
Db 1941 CTGGGTGTCATAAACCTGCTGTGGCACCCAGCACCCGCCTATCCCTATCTG 1782
Qy 251 TGGAAACAGAGACATCAAGGTGCCAAGAAAGGTGGTAAAGCCAGCGCTCTT 310
Db 1781 TGGAAACAGAGACATCAAGGTGCCAAGAAAGTGGTAAAGCCACAGCGCTCTT 1722
Qy 311 CAACCTGCAAGTCACATCAAGTCACATCCAGCTCAACTCTGCGGGGCTG 370
Db 1721 CAACCTGCAAGTCACATCAACTCAACTCCAGACCTCTACTCTGCTGGGCTC 1662

repeat_region	18891..19237 /rpt_family="AluSx"	Db	25356 CA 25355
repeat_region	19239..19387 /rpt_family="AluY"	RESULT 10 AC087645/c	AC087645 LOCUS Homo sapiens chromosome 17, clone RP1-219G17, complete sequence. DEFINITION Homo sapiens chromosome 17, clone RP1-219G17, complete sequence. ACCESSION AC087645 VERSION HTG.
repeat_region	19388..19675 /rpt_family="AluSc"		
repeat_region	19676..19680 complement(20303..20438) /rpt_family="L1M9"		
repeat_region	complement(20799..21088) /rpt_family="AluSp"		
repeat_region	complement(21177..21230) /rpt_family="MIR" complement(22718..23024) /rpt_family="AluJb"		
repeat_region	complement(23053..23147) /rpt_family="Alu53"		
repeat_region	complement(23349..23524) /rpt_family="AluJ0" 23525..23840		
repeat_region	/rpt_family="AluSx" complement(23821..23936)		
repeat_region	/rpt_family="AluO" complement(23947..24088)		
repeat_region	/rpt_family="MIR" 24216..24415 complement(24316..24615)		
repeat_region	/rpt_family="AluSx" 24616..24736		
repeat_region	/rpt_family="MIR" complement(24742..25021)		
repeat_region	/rpt_family="AluSx" 25728..26024		
repeat_region	/rpt_family="AluSx" complement(26044..26353)		
repeat_region	/rpt_family="AluSq" 26354..26416 /rpt_family="L1PA16"		
repeat_region	26458..27006 complement(27018..27252)		
repeat_region	/rpt_family="AluSq" complement(27256..27450)		
repeat_region	27451..27452 /rpt_family="AluSx"		
Query Match	30.3% Score 300.4; DB 9; Best Local Similarity 99.7%; Prod. No. 1e-61;		
Matches 301; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY 131 AGAAATTACCCCTGGCTCATATGCCAAGTCTGGAGATTCCTCTCG 190			
Db 25656 AGAAATTACCCCTGGCTCATATGCCAAGTCTGGAGATTCCTCTCG 25597			
QY 191 CTGGGTGCTCATACCTGCTGCACTGCTGCACTGCTGCACTGCTGCTG 250			
Db 25596 CTGGGTGCTCATACCTGCTGCACTGCTGCACTGCTGCTGCTG 25537			
QY 251 TGGAAACCAGAACATCAAGTGGCCAGAAGGTGGTGAAGACCCACAGC 310			
Db 25536 TGGAAACCAGAACATCAAGTGGCCAGAAGGTGGTGAAGACCCACAGC 25477			
QY 311 CAACTCTAACGTCAACTCAAGTCAGTGCTCACCTACTGCTGGGCCTGC 370			
Db 25476 CAACTCTAACGTCAACTCAAGTCAGTGCTCACCTACTGCTGGGCCTGC 25417			
QY 371 CTCACCTCTAACGTGGCCATGTCAGGTACAGTGACTCTGGCTGGTC 430			
Db 25416 CTACCTCTAACGTGGCCATGTCAGGTACAGTGACTCTGGCTGGTC 25357			
QY 431 CA 432			

Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kanat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
 McDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Medrini,J., Meneus,L., Mlenga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retska,R., Ribe,C., Rogov,P.,
 Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., TeFayev,J., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., VieJ,R., Vo,A., Wilson,B., Wu,X.,
 Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (05-FEB-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 20, 2002 this sequence version replaced gi:26190573.
 All repeats were identified using RepeatMasker:
 Smits,A.F.A. & Green,P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute / MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: I11992
 Center clone name: 219_G_17
 ----- Location/Qualifiers
 1..209751
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="RP11-219G17"
 /clone_lib="RPCI-11 Human Male BAC"
 repeat_region 95..232
 /rpt_family="C-rich"
 complement(634..972)
 repeat_region /rpt_family="AluSx"
 complement(1653..1760)
 repeat_region /rpt_family="MBS5A"
 complement(1807..1886)
 repeat_region /rpt_family="MIR"
 complement(2001..2312)
 repeat_region /rpt_family="AluJ0"
 complement(2316..2636)
 repeat_region /rpt_family="AluSx"
 complement(3684..3706)
 repeat_region /rpt_family="AluJ0"
 complement(3707..3993)
 repeat_region /rpt_family="AluSx"
 complement(3994..4267)
 repeat_region /rpt_family="AluJ0"
 complement(5194..5371)
 repeat_region /rpt_family="AluSg/x"
 complement(5606..5738)
 repeat_region /rpt_family="FRAM"
 complement(5739..5850)
 repeat_region /rpt_family="AluU"
 complement(5865..5999)
 repeat_region /rpt_family="AluSg/x"
 complement(6293..6600)
 repeat_region /rpt_family="Alusp"
 complement(7562..7879)
 repeat_region /rpt_family="Aluy"
 complement(8178..8336)
 repeat_region /rpt_family="LIME4a"
 complement(8327..8617)
 repeat_region /rpt_family="AluSx"
 complement(8618..8721)

/rpt_family="L1ME4a"
 repeat_region 8799..9231
 /rpt_family="L3"
 repeat_region 9248..9498
 /rpt_family="L3"
 repeat_region 9596..10097
 /rpt_family="AluSg/x"
 complement(10100..10397)
 /rpt_family="AluY"
 10159..10164
 /rpt_family="AluSg/x"
 unsure
 /note=<30 qual SNGL region<
 repeat_region 10399..10579
 /rpt_family="AluSg/x"
 repeat_region 10580..10598
 /rpt_family="AluSx"
 repeat_region 10909..10919
 /rpt_family="AluSg/x"
 repeat_region 11385..11522
 /rpt_family="MIR"
 complement(11696..11997)
 /rpt_family="Aluy"
 complement(12826..12967)
 repeat_region 10909..10919
 /rpt_family="L1MC4"
 complement(12968..13262)
 /rpt_family="AluJb"
 complement(13263..13344)
 /rpt_family="Aluy"
 complement(13354..13532)
 repeat_region 13345..13532
 /rpt_family="AluJb"
 complement(13533..13820)
 /rpt_family="AluSx"
 complement(13821..13890)
 /rpt_family="L1MC4"
 repeat_region 14283..14348
 /rpt_family="MER45C"
 complement(14349..14525)
 repeat_region 14349..14525
 /rpt_family="AluSg"
 complement(14526..14828)
 /rpt_family="Aluy"
 complement(14829..14962)
 repeat_region 14963..14999
 /rpt_family="AluSg"
 complement(14999..15299)
 repeat_region 15303..15558
 /rpt_family="AluSx"
 repeat_region 15599..15752
 /rpt_family="MIR"
 complement(15752..16198)
 repeat_region 15752..16198
 /rpt_family="MER45C"
 complement(16198..16332)
 repeat_region 16832..16961
 /rpt_family="MIR"
 complement(16961..17736)
 /rpt_family="MIR"
 unsure
 /note="single clone coverage"
 repeat_region 17736..17810
 /rpt_family="AluSg"
 complement(17810..17975)
 repeat_region 1794..17975
 /rpt_family="CT-rich"
 repeat_region 18235..18553
 /rpt_family="AluSg"

Query Match 30.3%; Score 300.4; DB 9; Length 209751;
 Best Local Similarity 90.7%; Pred. No. 1e-61; 1; Indels 0; Gaps 0;

	Matches 301; Conservative 0; Mismatches 0;
Qy	131 AGAAATTACCCCTGTCTCATAGTCTAACAGTCTGGAGTTTCCTCAAAGGCCG 190
Db	181172 AGAAATTACCCCTGTCTCATAGTCTAACAGTCTGGAGTTTCCTCAAAGGCCG 181113
Qy	191 CTGGGTGCTCATACCTGCTGTGACCCAGGCCATACCTATTCCTCTG 250
Db	181112 CTGGGTGCTCATACCTGCTGTGACCCAGGCCATACCTATTCCTCTG 181053
Qy	251 TGGAACAAAGAACATCAAGTGGCAAGAACGGTGTGAAGACCCAGAGCGGCCTCTT 310

Db 181052 TGGAAACAGAACATCAAGGTGCCAAGAAGGTGGTGAAGACCACAGCCGGCTCCTT 180993
 Qy 311 CAACCTCAAGTCACACTCAAGTCAGTCAGACGAGCTGCTACACTCTGGCGGGCCTC 370
 Db 180992 CAACTCTAACGTCACACTCAAGTCAGTCAGACGAGCTGCTACACTCTGGCGGGCCTC 180933
 Qy 371 CTGCCACCTCAGTGCCCATGTGAGACATGCAAGCTCAGCTGAGATGTTGTC 430
 Db 180932 CTACACCTCAGGCCATGTGAGACATGCAAGCTCAGCTGAGATGTTGTC 180873
 Qy 431 CA 432
 Db 180872 CA 180871

RESULT 11
 AC010532 AC010532 DNA 229426 bp LINEAR HTG 05-SEP-2000
 LOCUS Homo sapiens chromosome 17 cClone RP11-219G17, LOW-PASS SEQUENCE
 DEFINITION SAMPLING.
 AC010532 AC010532.2 GI:6758797
 VERSION HTGS_PHASE0
 KEYWORDS Homo sapiens (human)
 SOURCE Organism
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;
 Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 229426)
 Unpublished
 Sequencing of Human Chromosome 17
 REFERENCE Jan 26, 2000 This sequence version replaced gi:58B2406.
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 17
 JOURNAL
 REFERENCE 2 (bases 1 to 229426)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission.
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 COMMENT Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>
 NOTE: This record contains 99 individual sequencing reads that have not been assembled into
 contigs. Runs of N are used to separate the reads
 and the order in which they appear is completely
 arbitrary. Low-pass sequence sampling is useful for
 identifying clones that may be gene-rich and allows
 overlap relationships among clones to be deduced.
 However, it should not be assumed that this clone
 will be sequenced to completion. In the event that
 the record is updated, the accession number will
 be preserved.
 1 563: contig of 563 bp in length
 * 564 1436: contig of 873 bp in length
 * 1437 1649: contig of 213 bp in length
 * 1650 2523: contig of 874 bp in length
 * 2524 3286: contig of 763 bp in length
 * 3287 3613: contig of 327 bp in length
 * 3614 5430: contig of 1817 bp in length
 * 5431 6681: contig of 1251 bp in length
 * 6682 7357: contig of 676 bp in length
 * 7358 8117: contig of 760 bp in length

* 8118 8718: gap of unknown length
 * 8719 9464: contig of 601 bp in length
 * 9465 9964: contig of 746 bp in length
 * 9965 10959: contig of 500 bp in length
 * 9965 10959: contig of 995 bp in length
 * 10950 11873: contig of 914 bp in length
 * 10950 11873: gap of unknown length
 * 11874 12656: contig of 914 bp in length
 * 11874 12656: gap of unknown length
 * 12657 13674: contig of 1018 bp in length
 * 12657 13675: gap of unknown length
 * 13675 14935: contig of 1261 bp in length
 * 14936 16423: contig of 763 bp in length
 * 16424 17186: contig of 1488 bp in length
 * 17187 18079: gap of unknown length
 * 18080 18813: contig of 893 bp in length
 * 18814 19729: contig of 934 bp in length
 * 18814 19729: gap of unknown length
 * 19730 20675: contig of 946 bp in length
 * 20676 21721: gap of unknown length
 * 21722 23111: contig of 1046 bp in length
 * 21722 23111: gap of unknown length
 * 23112 25188: contig of 1390 bp in length
 * 23112 25188: gap of unknown length
 * 25189 26588: contig of 2077 bp in length
 * 25189 26588: gap of unknown length
 * 26589 27644: contig of 1400 bp in length
 * 26589 27644: gap of unknown length
 * 27645 29372: contig of 1056 bp in length
 * 27645 29372: gap of unknown length
 * 29373 29495: contig of 1728 bp in length
 * 29373 29495: gap of unknown length
 * 29496 30431: contig of 123 bp in length
 * 29496 30431: gap of unknown length
 * 30431 31440: contig of 936 bp in length
 * 30431 31440: gap of unknown length
 * 31450 32983: contig of 1018 bp in length
 * 31450 32983: gap of unknown length
 * 32994 33162: contig of 1534 bp in length
 * 32994 33162: gap of unknown length
 * 33163 33975: contig of 179 bp in length
 * 33163 33975: gap of unknown length
 * 33976 35031: contig of 1056 bp in length
 * 33976 35031: gap of unknown length
 * 35032 35709: contig of 678 bp in length
 * 35032 35709: gap of unknown length
 * 35710 37308: contig of 1599 bp in length
 * 35710 37308: gap of unknown length
 * 37309 38107: contig of 799 bp in length
 * 37309 38107: gap of unknown length
 * 38108 39102: contig of 995 bp in length
 * 38108 39102: gap of unknown length
 * 39103 39591: contig of 489 bp in length
 * 39103 39591: gap of unknown length
 * 39592 41151: contig of 1560 bp in length
 * 39592 41151: gap of unknown length
 * 41152 42842: contig of 1691 bp in length
 * 41152 42842: gap of unknown length
 * 42843 43888: contig of 698 bp in length
 * 42843 43888: gap of unknown length
 * 44000: contig of 181 bp in length
 * 43839 44000: contig of 181 bp in length
 * 43839 44000: gap of unknown length

*	44010	44874: contig of 865 bp in length	29.5%; Score 292.8; DB 2; Length 229426;
**	44875	gap of unknown length	Best Local Similarity 97.7%; Pred. No. 6.BE-60; Mismatches 0; Conservative Matches 297; Indels 0; Gaps 0;
**	45191:	contig of 317 bp in length	
**	46147:	gap of unknown length	
**	46148	contig of 956 bp in length	688 CAGTGTTGCCCAAGTGGTACCGAAATGGGACTCTGGAGCTGGAGCTGGAGCC 747
**	47504:	gap of unknown length	Db 135296 CACIGCTTGTCAAGGTGGTACCGAAATGGGACTGGAGCTGGAGCC 135355
**	49011:	contig of 1357 bp in length	Qy 748 CCATCCTTGCCTTCGGCTTACAGGAGCACCCGGCTGAGTGAAGGAGTTGGGG 807
**	50259:	gap of unknown length	Db 135356 CCATCCTTGCCTTCGGCTTACAGGAGCACCCGGCTGAGTGAAGGAGTTGGGG 135415
**	50260	contig of 1168 bp in length	Qy 808 GGTTCA GGATTGGAAATGGGGAGGTCAAGGAGCTAACAGCAGCAAGTAGATAAAC 867
**	51361:	gap of unknown length	Db 135416 GGTTAGGATGGGAATGGGGAGGTCAAGGAGCTAACAGCAGCAAGTAGATAAAC 135475
**	51362	contig of 1270 bp in length	
**	52632	gap of unknown length	Qy 868 CGTCCAGAAGGCCAAGCAGGAGACTGCGAGCATAGCGTGCAGTGTGACTGTGCTATT 927
**	53608	contig of 1213 bp in length	Db 135476 CGTCCA GGAGGCCAAGCAGGAGACTGCGAGCATAGCGTGCAGTGTGACTGTGCTATT 135535
**	54820:	gap of unknown length	
**	56600:	contig of 1780 bp in length	Qy 928 GGAGTTCATGCCAAATGAGTGTGTTAGTGTGCTCTGCCACAAAAAAAGAAAAAAA 987
**	58236:	gap of unknown length	Db 135536 GGAGTTCATGCCAAATGAGTGTGTTAGTGTGCTCTGCCACAAAAAAAGAAAAAAA 135595
**	58237	contig of 1480 bp in length	Qy 988 AAAA 991
**	59717	gap of unknown length	Db 135596 AAAA 135599
<hr/>			
RESULT 12			
AC032035 LOCUS 154840 bp DNA linear HTG 08-SPB-2000			
DEFINITION Homo sapiens chromosome 17 clone RP11-141D15 map 17, *** SEQUENCING			
AC032035 IN PROGRESS **, 18 unordered pieces.			
ACCESSION AC032035_3 GI:9994161			
VERSION AC032035_3 GI:9994161			
KEYWORDS HTG; HTG PHASEI.			
ORGANISM Homo sapiens (human)			
SOURCE Homo sapiens			
REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.			
TITLE Homo sapiens chromosome 17, clone RP11-141D15			
JOURNAL Unpublished			
REFERENCE 2 (bases 1 to 154840)			
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,			
Boguslavsky,L., Boukhalter,A., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardya,S., Ginde,S., Goette,M., Graham,L., Grand,Pierre,N., Grafit,G., Hagos,B., Heatord,A., Horton,L., Howland,J.C., Ilie,I., Johnson,R., Jones,C., Kann,I., Karatas,A., Klein,J., LaRocque,X., Lanazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,D., Marquis,N., McCarthy,M., McEwan,P., McGrath,A., McKernan,K., McPheeeters,R., McEldrim,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,J., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,			
COMMENT Direct Submission			
Submitted (03-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
On Sep 8, 2000 this sequence version replaced gi:7705196.			
All repeats were identified using RepeatMasker:			
Smit, A.F.A. & Green, P. (1996-1997)			

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Genome Center

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence submission@genome.wi.mit.edu

Project Information

Center Project name: L9138

Center Clone name: 141_D_15

* NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2373: contig of 2373 bp in length

2473: gap of 100 bp

4930: contig of 2457 bp in length

5030: gap of 100 bp

7215: contig of 2185 bp in length

7315: gap of 100 bp

7316: contig of 2226 bp in length

9542: gap of 100 bp

9641: gap of 100 bp

9642: 12211: contig of 2570 bp in length

12311: gap of 100 bp

12312: contig of 2612 bp in length

14923: gap of 100 bp

15023: gap of 100 bp

15024: 19094: contig of 4011 bp in length

19095: 19195: gap of 100 bp

22286: contig of 3092 bp in length

22287: 23386: gap of 100 bp

22387: 26147: contig of 3161 bp in length

26148: 26247: gap of 100 bp

26248: 31933: contig of 5886 bp in length

31934: 32033: gap of 100 bp

32034: 37339: contig of 5306 bp in length

37340: 37439: gap of 100 bp

37440: 42858: contig of 5419 bp in length

42859: 42959: gap of 100 bp

42959: 44810: contig of 5552 bp in length

44811: 48910: gap of 100 bp

48911: 56029: contig of 7119 bp in length

56030: 56129: gap of 100 bp

56130: 62189: contig of 6680 bp in length

62290: 62289: gap of 100 bp

62290: 82929: contig of 6670 bp in length

82929: 83059: gap of 100 bp

83059: 83059: contig of 24000 bp in length

107060: 107159: gap of 100 bp

107160: 154840: contig of 47681 bp in length.

Location/Qualifiers

1. 154840

/organism="Homo sapiens"

/mol_type="Genomic DNA"

/db_xref="taxon:9606"

/chromosome="17"

/clone="P11-141D15"

/clone_lib="RP01-11 Human Male BAC"

36180 a 41741 c 41557 g 32874 t 2488 others

FEATURES
source

BASE COUNT ORIGIN

Query Match Score 291.2; DB 2; Length 154840;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE COUNT ORIGIN

Query Match Score 29.4%;

Best Local Similarity 97.4%; Pred. No. 1.7e-39;

Matches 296; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

FEATUR

source

BASE

*	5431:	6681: contig of 1251 bp in length	gap of unknown length
*	6682:	gap of unknown length	contig of 986 bp in length
*	7357:	contig of 676 bp in length	gap of unknown length
*	7358:	gap of unknown length	contig of 181 bp in length
*	8117:	contig of 760 bp in length	gap of unknown length
*	8718:	gap of unknown length	contig of 865 bp in length
*	8718:	contig of 601 bp in length	gap of unknown length
*	8719:	gap of unknown length	contig of 317 bp in length
*	9464:	contig of 746 bp in length	gap of unknown length
*	9964:	gap of unknown length	contig of 956 bp in length
*	10959:	contig of 500 bp in length	gap of unknown length
*	9965:	gap of unknown length	contig of 1357 bp in length
*	10960:	contig of 995 bp in length	gap of unknown length
*	11873:	contig of 914 bp in length	contig of 1587 bp in length
*	12656:	gap of unknown length	contig of 956 bp in length
*	13674:	contig of 1018 bp in length	gap of unknown length
*	13675:	contig of 1261 bp in length	contig of 1102 bp in length
*	14935:	gap of unknown length	contig of 1102 bp in length
*	16423:	contig of 1488 bp in length	gap of unknown length
*	17186:	contig of 763 bp in length	contig of 1213 bp in length
*	17187:	gap of unknown length	contig of 1270 bp in length
*	18079:	contig of 893 bp in length	gap of unknown length
*	18813:	contig of 734 bp in length	contig of 1168 bp in length
*	18814:	gap of unknown length	contig of 1168 bp in length
*	19730:	contig of 916 bp in length	gap of unknown length
*	20675:	contig of 946 bp in length	contig of 1270 bp in length
*	21721:	contig of 1146 bp in length	gap of unknown length
*	23111:	contig of 1330 bp in length	contig of 1270 bp in length
*	23112:	gap of unknown length	gap of unknown length
*	25188:	contig of 2077 bp in length	contig of 1270 bp in length
*	26588:	contig of 1400 bp in length	gap of unknown length
*	26589:	gap of unknown length	contig of 1270 bp in length
*	27644:	contig of 1056 bp in length	gap of unknown length
*	29372:	contig of 1728 bp in length	contig of 1270 bp in length
*	29495:	contig of 123 bp in length	gap of unknown length
*	29496:	contig of 936 bp in length	contig of 1270 bp in length
*	30432:	gap of unknown length	gap of unknown length
*	32983:	contig of 1018 bp in length	contig of 1270 bp in length
*	32984:	gap of unknown length	gap of unknown length
*	33163:	contig of 813 bp in length	contig of 1270 bp in length
*	33976:	gap of unknown length	gap of unknown length
*	35031:	contig of 1534 bp in length	contig of 1270 bp in length
*	35032:	gap of unknown length	gap of unknown length
*	35709:	contig of 678 bp in length	contig of 1270 bp in length
*	33975:	gap of unknown length	gap of unknown length
*	37308:	contig of 1599 bp in length	contig of 1270 bp in length
*	38107:	contig of 799 bp in length	gap of unknown length
*	39102:	contig of 95 bp in length	contig of 1270 bp in length
*	39103:	contig of 489 bp in length	gap of unknown length
*	41151:	contig of 1560 bp in length	contig of 1270 bp in length
*	39592:	gap of unknown length	gap of unknown length
*	41152:	contig of 1691 bp in length	gap of unknown length

		FEATURES	SOURCE	Location/Qualifiers
*	111098	114888: contig of 3791 bp in length		1. -116300 /organism="Mus musculus"
*	114889	117423: contig of unknown length		/mol_type="Genomic DNA" /db_xref="taxon:10090" /chromosome="11" /clone="RP23-151N19" /clone_lib="RPCI-23"
		gap of unknown length.		
Query Match		Score 179.4; DB 2; Length 229426;		
Best Local Similarity	18.1%	Pred. No. 1.5e-32;		
Matches	183;	Mismatches 0; Gaps 0;		
Db	1331	CGCTGGGAAACCAAGAACATCAGGTGCCAGAAGTGGTAAGCCACAGGCCG 303		BASE COUNT 29288 a 28679 c 29788 g 28545 t
Qy	244	CCCTCTGTGGACCAAGAACATCAGGTGCCAGAAGTGGTAAGCCACAGGCCG 303		ORIGIN
Db	1271	CCTCTTACCTAACGTCACATCAAGTCCAGTCCAGTCTCCAGAATTCCTGCT 363		Query Match 16.5%; Score 163.4; DB 10; Length 116300;
Qy	304	CCTCTTCAACTAACGTCACATCAAGTCCAGTCTCCAGAATTCCTGCT 363		Best Local Similarity 73.3%; Pred. No. 1.1e-26;
Db	1271	CCTCTTACCTAACGTCACATCAAGTCCAGTCTCCAGAATTCCTGCT 1212		Matches 209; Conservative 0; Mismatches 76; Gaps 0;
Qy	364	GGGGGTCCCTCACCTTCAGGTGCCCATGCGACAGTGGTCAAGTGAATGGAGC 423		Query Match 14.8%; Score 143.4; DB 10; Length 116300;
Db	1211	GGGGTCCCTCACCTTCAGTCAGTGACAGTGGTCAAGTGAATGGAGC 1152		Best Local Similarity 73.3%; Pred. No. 1.1e-26;
Qy	424	TGTTGTCCA 432		Matches 209; Conservative 0; Mismatches 76; Gaps 0;
Db	1151	TGTTGTCCA 1143		Query Match 14.8%; Score 143.4; DB 10; Length 116300;
RESULT 14				Best Local Similarity 73.3%; Pred. No. 1.1e-26;
AL645856	AL645856	AL645856 AL645856 AL645856 AL645856 AL645856 AL645856		Matches 209; Conservative 0; Mismatches 76; Gaps 0;
DEFINITION	Mouse DNA sequence from clone RP23-151N19	linear ROD 24-AUG-2002		Query Match 14.8%; Score 143.4; DB 10; Length 116300;
COMMENT	complete sequence.	on chromosome 11,		Best Local Similarity 73.3%; Pred. No. 1.1e-26;
ACCESSION	AL645856	GI:22531400		Matches 209; Conservative 0; Mismatches 76; Gaps 0;
VERSION	AL645856.5	HTG.		Query Match 14.8%; Score 143.4; DB 10; Length 116300;
KEYWORDS				Best Local Similarity 73.3%; Pred. No. 1.1e-26;
SOURCE				Matches 209; Conservative 0; Mismatches 76; Gaps 0;
ORGANISM		Mus musculus (house mouse)		Query Match 14.8%; Score 143.4; DB 10; Length 116300;
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				Best Local Similarity 73.3%; Pred. No. 1.1e-26;
1. (bases 1 to 116300)				Matches 209; Conservative 0; Mismatches 76; Gaps 0;
REFERENCE	AL954690	AL954690 AL954690 AL954690 AL954690 AL954690 AL954690		Query Match 14.8%; Score 143.4; DB 10; Length 116300;
AUTHORS	White,S.	LOCUS	AL954690	Best Local Similarity 73.3%; Pred. No. 1.1e-26;
TITLE		DEFINITION	AL954690	Matches 209; Conservative 0; Mismatches 76; Gaps 0;
JOURNAL		PROGRESS ***	AL954690	Query Match 14.8%; Score 143.4; DB 10; Length 116300;
Submitted (24-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		ACCESSION	AL954690.2	Best Local Similarity 73.3%; Pred. No. 1.1e-26;
On Aug 27, 2002 this sequence replaced gi:21738435.		VERSION	GT:25136851	Matches 209; Conservative 0; Mismatches 76; Gaps 0;
----- Genome Center		KEYWORDS	HTGS; PHASE2; HTGS_CANCELLED.	Query Match 14.8%; Score 143.4; DB 10; Length 116300;
Center code: SC		SOURCE	Mus musculus	Best Local Similarity 73.3%; Pred. No. 1.1e-26;
Web site: http://www.sanger.ac.uk		ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	Matches 209; Conservative 0; Mismatches 76; Gaps 0;
Contact: humquery@sanger.ac.uk		REFERENCE	1 (bases 1 to 197194)	Query Match 14.8%; Score 143.4; DB 10; Length 116300;
-----		AUTHORS	Tromans,A.	Best Local Similarity 73.3%; Pred. No. 1.1e-26;
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.		TITLE	Direct Submission	Matches 209; Conservative 0; Mismatches 76; Gaps 0;
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL:, Sw:, SWISSPROT:, Tr:, TREMBL;, Wp:, WORMBEP;. Information on the WORMBEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep/		JOURNAL	Submitted (13-NOV-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	Query Match 14.8%; Score 143.4; DB 10; Length 116300;
COMMENT		COMMENT	On Nov 19, 2002 this sequence version replaced gi:25045706.	Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
			-----	----- Genome Center
				Center: Wellcome Trust Sanger Institute
				Center code: SC
				Web site: http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
				Contact: humquery@sanger.ac.uk
				----- Project Information
				Center project name: bM38644
				----- Summary Statistics
				Assembly program: XGA4; version 4.5
				Chemistry: Dye-terminator; www.sanger.ac.uk
				Consensus quality: 195990 bases at least Q40
				Consensus quality: 196234 bases at least Q30
				Consensus quality: 196338 bases at least Q20
				Insert size: 197194; sum-of-contigs
				VECTO: PBACE3.6.

Lamazares, R., Landers, T., Lebockzy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, C., McEwan, P., McKernan, K., McGeheeers, R., Meldrum, J., Menus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Conor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Polaria, V., Raymond, C., Retta, R., Rieley, R., Rose, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Severy, P., Spence, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testayre, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliou, H., Visel, R., Wong, A., Wilson, D., Wu, X., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

Submitted: (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:
Shiu, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contract: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L15700
Center clone name: 154_K_21

* NOTE: This record contains 48 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

*	1	986: contig of 986 bp in length
*	987	1086: gap of 100 bp
*	1087	2009: contig of 923 bp in length
*	2010	2109: gap of 100 bp
*	2110	3007: contig of 898 bp in length
*	3008	3107: gap of 100 bp
*	3108	3866: contig of 759 bp in length
*	3867	3966: gap of 100 bp
*	3967	4874: contig of 908 bp in length
*	4875	4974: gap of 100 bp
*	4975	5839: contig of 865 bp in length
*	5840	5939: gap of 100 bp
*	5940	7017: contig of 1078 bp in length
*	7018	7117: gap of 100 bp
*	7118	8079: contig of 962 bp in length
*	8080	8179: gap of 100 bp
*	8180	9139: contig of 960 bp in length
*	9140	9239: gap of 100 bp
*	9240	10154: contig of 1015 bp in length
*	10255	10354: gap of 100 bp
*	10355	11289: contig of 935 bp in length
*	11290	11389: gap of 100 bp
*	11390	12334: contig of 945 bp in length
*	12335	12434: gap of 100 bp
*	12435	13410: contig of 976 bp in length
*	13411	13510: gap of 100 bp
*	13511	14445: contig of 935 bp in length
*	14446	14545: gap of 100 bp
*	14546	15272: contig of 727 bp in length
*	15273	15372: gap of 100 bp
*	15373	16394: contig of 1022 bp in length
*	16395	16494: gap of 100 bp

FEATURES	source	Location/Qualifiers	1.4932 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /clone_id="R223-154K21"	16.1%; Score 160.2; DB 2; Bred. No. 6.6e-28;
17562	*	18545: contig of 984 bp in length		
18645	*	18645: gap of 100 bp		
18646	*	18646: contig of 955 bp in length		
19600	*	19600: gap of 100 bp		
19700	*	19700: contig of 970 bp in length		
20670	*	20770: gap of 100 bp		
20771	*	21739: contig of 969 bp in length		
21740	*	21839: gap of 100 bp		
21840	*	22669: contig of 830 bp in length		
22670	*	22769: gap of 100 bp		
23763	*	23763: contig of 994 bp in length		
23863	*	23863: gap of 100 bp		
23864	*	24820: contig of 957 bp in length		
24821	*	24920: gap of 100 bp		
25847	*	25847: contig of 927 bp in length		
25947	*	25947: gap of 100 bp		
26787	*	26787: contig of 840 bp in length		
26888	*	26888: gap of 100 bp		
27831	*	27831: contig of 944 bp in length		
27932	*	27932: gap of 100 bp		
28882	*	28882: 28981: gap of 100 bp		
28982	*	29906: contig of 925 bp in length		
29907	*	30006: gap of 100 bp		
30007	*	30921: contig of 915 bp in length		
30922	*	31021: gap of 100 bp		
31022	*	31757: contig of 736 bp in length		
31758	*	31857: gap of 100 bp		
31858	*	32798: contig of 941 bp in length		
32799	*	32898: gap of 100 bp		
32899	*	33835: contig of 937 bp in length		
33836	*	33935: gap of 100 bp		
33936	*	34960: contig of 925 bp in length		
34961	*	34961: gap of 100 bp		
35888	*	35887: contig of 927 bp in length		
35889	*	35889: gap of 100 bp		
36990	*	36990: contig of 1003 bp in length		
36991	*	37090: gap of 100 bp		
38085	*	38085: contig of 995 bp in length		
38086	*	38185: gap of 100 bp		
38186	*	39172: contig of 987 bp in length		
39173	*	39272: gap of 100 bp		
39273	*	40215: contig of 943 bp in length		
40216	*	40315: gap of 100 bp		
41267	*	41267: contig of 952 bp in length		
41268	*	41367: gap of 100 bp		
41368	*	42271: contig of 904 bp in length		
42272	*	42371: gap of 100 bp		
42372	*	43302: contig of 931 bp in length		
43303	*	43402: gap of 100 bp		
43403	*	44283: contig of 881 bp in length		
44284	*	44284: gap of 100 bp		
44384	*	45395: contig of 1012 bp in length		
45396	*	45495: gap of 100 bp		
45396	*	46390: contig of 895 bp in length		
46391	*	46490: gap of 100 bp		
46490	*	47408: contig of 918 bp in length		
47409	*	47508: gap of 100 bp		
47509	*	48438: contig of 930 bp in length		
48439	*	48538: gap of 100 bp		
48539	*	49532: contig of 994 bp in length.		
FEATURES	source	Location/Qualifiers	1.4932 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /clone_id="R223-154K21"	16.1%; Score 160.2; DB 2; Bred. No. 6.6e-28;
BASE COUNT	11202 a	11441 c	10738 g	11027 t
ORIGIN	Query Match	Best Local Similarity	5124 others	

FEATURES	source	Annealing: 58 degrees C for 1.50 minutes Polymerization: 72 degrees C for 1.00 minutes PCR Cycles: 30 Protocol: Thermal Cycler: custom built by IAS, Costar, Cambridge MA
STS	primer_bind primer_bind_complement(132..150) BASE COUNT 30 a 46 c 33 g 40 t 1 others ORIGIN	Template: 10 ng Primer: each 5 PM dNTPs: 4 uM Taq Polymerase: 0.5 uU Total Vol.: 20 uL
Buffer:		Mg2+: 1.5 mM KCl: 50 mM Tris-HCl: 10 mM Gelatin: .001 % Location/Qualifiers
		1. 1.150 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone.lib="Human Hudson EST" /note="STSs derived from sequences in dbEST and the UniGene collection." 1. .150 1. 20 primer_bind complement(132..150)
Query Match		Score 149; DB 11; Length 150; Best Local Similarity 99.3%; Pred. No. 3.8e-25; Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	820 GGAATGGGGAGGTAGGGCAAGCACGCCATGAAATGAACTGCGTATTGGAGTTCACTGC 879	/mol_type="mRNA" /db_xref="taxon:9606" /clone.lib="Human Hudson EST" /note="STSs derived from sequences in dbEST and the UniGene collection."
Db	150 GGAATGGGGAGGTAGGGCAAGCACGCCATGAAATGAACTGCGTATTGGAGTTCACTGC 91	1. 150 prime_bind 1. 20 primer_bind_complement(132..150)
Query Match		Score 149; DB 11; Length 150; Best Local Similarity 99.3%; Pred. No. 3.8e-25; Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	880 CAAGCACGGAGGACTGGAGGACTGGAGGATCAACNTGACTGTCGTTAGTGAGTCATGCA 919	STS BASE COUNT 30 a 47 c 32 g 40 t 1 others ORIGIN
Db	90 CAAGCACGGAGGACTGGAGGACTGGAGGATCAACNTGACTGTCGTTAGTGAGTCATGCA 31	Query Match 820 GGAATGGGGAGGTAGGGCAAGCACGCCATGAACTGCGTATTGGAGTTCACTGC 879 Best Local Similarity 99.3%; Pred. No. 3.8e-25; Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	940 AAATGAGTGTGTTAGCTGCTCTGCAAC 969	Db 150 GGAATGGGGAGGTAGGGCAAGCACGCCATGAACTGCGTATTGGAGTTCACTGC 91
Db	30 AAATGAGTGTGTTAGCTGCTCTGCAAC 1	Qy 880 CAAGCACGGAGGACTGGAGGATCAACNTGACTGTCGTTAGTGAGTCATGCA 31 Db 90 CAAGCACGGAGGACTGGAGGATCAACNTGACTGTCGTTAGTGAGTCATGCA 31
RESULT 18		Query Match 940 AAATGAGTGTGTTAGCTGCTCTGCAAC 969 Best Local Similarity 99.3%; Pred. No. 3.8e-25; Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
LOCUS	G43546	STS 28-JAN-1999
DEFINITION	WIAF-2407-STS Human Hudson EST Homo sapiens STS cDNA, sequence tagged site.	LOCUS AC143361 161043 bp DNA linear HTG 09-APR-2003 DEFINITION Macaca mulatta clone CH20-271N12, *** SEQUENCING IN PROGRESS ***.
ACCESSION	G43546	ACCESSION AC143361 1 GI:29648548 VERSION G43546..1 GI:4192463 KEYWORDS HTG; HTGS PHASII; HTGS PGI.
VERSION	G43546..1	SOURCE Macaca mulatta (rheesus monkey)
KEYWORDS	STS.	ORGANISM Macaca mulatta
SOURCE	Homo sapiens (human)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ORGANISM	Homo sapiens	RESULT 19 AC143361 AC143361 1 (bases 1 to 150) (bases 1 to 150) (bases 1 to 150)
REFERENCE	AUTHORS Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sabolsky, R., Chandour, G., Perkins, N.M., Winchester, E., Spencer, J., Kriglyak, L., Stein, L.L., Hsie, L., Topaloglu, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Liphshitz, R., Chee, M., and Lander, E.S.	DEFINITION Macaca mulatta clone CH20-271N12, *** SEQUENCING IN PROGRESS ***. ACCESSION AC143361 1 GI:29648548 VERSION HTG; HTGS PHASII; HTGS PGI.
COMMENT	TITLE Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome Science 280 (5366), 1077-1082 (1998)	SOURCE Macaca mulatta ORGANISM Macaca mulatta (rheesus monkey)
JOURNAL	98248615	JOURNAL (in) Guido, R. and Gusfield, D. (Eds.); Springer (2002) Algorithms in Bioinformatics, Second International Workshop, WABI 2002, Rome, Italy, September 17-21, 2002, Proceedings: 10-28; (bases 1 to 161043)
PUBMED	9582121	REFERENCE AUTHORS 1 (bases 1 to 161043) TITLE Csurowski, M. and Milosavljevic, A. Pooled genomic indexing (PGI). design JOURNAL (in) Guido, R. and Gusfield, D. (Eds.); Springer (2002) Algorithms in Bioinformatics, Second International Workshop, WABI 2002, Rome, Italy, September 17-21, 2002, Proceedings: 10-28; (bases 1 to 161043)
COMMENT	Synonyms: EST376246b, EST376246 Contact: Thomas Hudson Whitehead Institute/MIT Center for Genome Research Whitehead Institute for Biomedical Research 9 Cambridge Center, Cambridge MA 02142 USA Tel: 617 252 1900 Fax: 617 252 1902 Email: thudson@genome.wi.mit.edu	REFERENCE AUTHORS 2 (bases 1 to 161043) TITLE Milosavljevic, A.; Sodergren, E.; Csurowski, M.; Li, B.; Jackson, A.R.; Adams, C.; Adio-Oduola, B.; Ali-Osman, F.R.; Allen, C.; Alsbrooks, S.L.; Amaralung, H.C.; Are, J.R.; Ayrele, M.; Banks, T.; Barberia, J.; Benten, J.; Binage, K.; Blankenburg, K.; Bonnici, D.; Bouck, J.; Bowie, S.; Brieva, M.; Brown, E.; Brown, M.; Bryant, N.P.; Buhay, C., Denaturation: 94 degrees C for 4.00 minutes PCR Profile: 94 degrees C for 50.0 seconds

AUTHORS Worley K.C.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUL-1999) Human Genome Sequencing Center, Department
 Of Molecular and Human Genetics, Baylor College of Medicine, One
 6 (bases 1 to 151750)
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE Worley, K.C.

AUTHORS Direct Submission
 JOURNAL Submitted (07-AUG-1999) Human Genome Sequencing Center, Department
 Of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT Sequencing is completed to a minimum standard of double strand
 coverage with a minimum of 2 clones and 2 reads with no ambiguities
 or 2 chemistries with a minimum of 2 clones and 3 reads with no
 ambiguities. If the sequence quality does not meet this standard,
 it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
 Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
 Zhang.

Exon/Intron boundaries of identified genes were chosen if there
 were canonical splice junctions that maintained sequence continuity
 across the splice junctions.

Location.Qualifiers

1. -151750
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xp22"
 /clone_id="GS-510109, U19987, U209F2"
 /clone.lib="Genome Systems Human BAC library, Lawrence
 Livermore X chromosome library"

misc_feature

repeat_region
 complement(2773_2863)
 /note="Overlaps bases 166883 to 167256 in AC008008"
 /function="Overlap with clone AC008008"
 complement(2657_2689)
 /rpt_family="AT rich"
 complement(3472_3481)
 /rpt_family="MER5B"
 /rpt_family="AluSx"
 5843_6143
 /rpt_family="AT rich"
 complement(6889_7493)
 /rpt_family="LIMA3"
 /rpt_family="MER20"
 11128_111201
 /rpt_family="GGAA"
 111487_111618
 /standard_name="hz-1"
 111757_111981
 /rpt_family="MER20"
 12660_12782
 /rpt_family="L12"
 12879_13195
 /rpt_family="LMT1E"
 /rpt_family="AluJc"
 complement(15636_16099)
 complement(15636_16099)
 /rpt_family="MBR20"
 complement(16717_17085)
 /rpt_family="L1MB7"
 complement(17159_17374)
 /rpt_family="L1MB7"
 complement(174721_19544)
 /rpt_family="LIM4"
 complement(19654_19760)
 /rpt_family="L1MB7"

repeat_region
 complement(20189_20669)
 /rpt_family="L1MB7"
 /rpt_family="MLTFP"
 complement(21172_21792)
 /rpt_family="L1MB7"
 22005_2430
 /rpt_family="L1MCI"
 23332_23445
 /rpt_family="L1PA4"
 23469_23588
 /rpt_family="FLAM_A"
 complement(25042_25067)
 /rpt_family="POLY_A"
 complement(25163_25265)
 /rpt_family="LIPAS"
 25164_25431
 /rpt_family="MER31A"
 complement(25569_26063)
 /rpt_family="MER31A"
 complement(27469_27570)
 /rpt_family="MIR"
 28608_28613
 /rpt_family="(TAAA)n"
 complement(28713_29010)
 /rpt_family="AluDb"
 29312_29310
 /rpt_family="MIR"
 29312_29613
 /rpt_family="AluSx"
 30127_30683
 /rpt_family="MLTD"
 31072_31160
 /rpt_family="GAA)n"
 31161_31279
 /rpt_family="GGA)n"
 complement(31772_31891)
 /rpt_family="GA)n"
 32352_32572
 /rpt_family="MIR"
 complement(33455_33479)
 /rpt_family="AT rich"
 34912_35382
 /rpt_family="L1MA1"
 35405_35701
 /rpt_family="AluSg"
 35707_36176
 /rpt_family="L1MA1"
 36178_38428
 /rpt_family="L1PA2"
 STS 36698_37069
 /standard_name="L1MA1"
 Homo sapiens
 /db_xref="dbSTS:32826"
 38436_38566
 /rpt_family="L1MA1"
 38597_38795
 /rpt_family="L1PB3"
 complement(38769_40304)
 /rpt_family="L1PA12"
 40309_40669
 /rpt_family="AluY"
 complement(40626_40905)
 /rpt_family="L1PA12"
 40904_41028
 /rpt_family="L1PB3"
 complement(41977_42037)
 /rpt_family="AT rich"
 complement(42391_42718)
 /rpt_family="MER34"

repeat_region
 complement(21171_21171)
 /rpt_family="MLTFP"
 complement(22005_22750)
 /rpt_family="L1MB7"
 23332_23445
 /rpt_family="L1PA4"
 23469_23588
 /rpt_family="FLAM_A"
 complement(25042_25067)
 /rpt_family="POLY_A"
 complement(25163_25265)
 /rpt_family="LIPAS"
 25164_25431
 /rpt_family="MER31A"
 complement(25569_26063)
 /rpt_family="MER31A"
 complement(27469_27570)
 /rpt_family="MIR"
 28608_28613
 /rpt_family="(TAAA)n"
 complement(28713_29010)
 /rpt_family="AluDb"
 29312_29310
 /rpt_family="MIR"
 29312_29613
 /rpt_family="AluSx"
 30127_30683
 /rpt_family="MLTD"
 31072_31160
 /rpt_family="GAA)n"
 31161_31279
 /rpt_family="GGA)n"
 complement(31772_31891)
 /rpt_family="GA)n"
 32352_32572
 /rpt_family="MIR"
 complement(33455_33479)
 /rpt_family="AT rich"
 34912_35382
 /rpt_family="L1MA1"
 35405_35701
 /rpt_family="AluSg"
 35707_36176
 /rpt_family="L1MA1"
 36178_38428
 /rpt_family="L1PA2"
 STS 36698_37069
 /standard_name="L1MA1"
 Homo sapiens
 /db_xref="dbSTS:7821"
 37958_38059
 /standard_name="A002D07", Chr. -, Homo sapiens

repeat_region
 complement(6889_7493)
 /rpt_family="LIMA3"
 complement(7513_7811)
 /rpt_family="AluSx"
 11128_111201
 /rpt_family="GGAA"
 111487_111618
 /standard_name="hz-1"
 111757_111981
 /rpt_family="MER20"
 12660_12782
 /rpt_family="L12"
 12879_13195
 /rpt_family="LMT1E"
 complement(15634_15659)
 /rpt_family="AluJc"
 complement(15636_16099)
 complement(16717_17085)
 /rpt_family="L1MB7"
 complement(17159_17374)
 /rpt_family="L1MB7"
 complement(174721_19544)
 /rpt_family="LIM4"
 complement(19654_19760)
 /rpt_family="L1MB7"

```

repeat_region          42979   . 43067
repeat_region          /rpt_Family="L2"
repeat_region          complement(43085
repeat_region          /rpt_Family="MIR"
repeat_region          43446   . 43746
repeat_region          /rpt_Family="AlusX"
repeat_region          44344   . 44380
repeat_region          /rpt_Family="(CAAA)n"
repeat_region          complement(45352
repeat_region          /rpt_Family="MTR"
repeat_region          complement(46643
repeat_region          /rpt_Family="L2"
repeat_region          complement(46905
repeat_region          /rpt_Family="MSTA"

Query Match           7.0%; Score 69.2; DB 9; Length 151750;
Best Local Similarity 66.8%; Pred. No. 5.7e-06;
Matches 131; Conservative 0; Mismatches 58; Indels 7; Gaps 2;
Matched bases: 796 AGGAGTTGGGGNTTCAGGATAGGGATCGGGACGGTCAGAGGCCAAGGGAGGCCA 855
Db      5846 AGGAAGTAAACCAGGGGTGGAAATAGGGAGATGGGGATAGGAGAGGATAAGAGGAGTA 5905
Dy      856 TCTAGAAATGAAACCTTCAGAGGCCAAGGCCA - -- GAGGACTCAGGCCATCGCGT 912
Db      5906 TGTTAGGATGAACTAACATTCTGGAGATCTAACACCCACTGGAGCTAATGGCATAAATT 5965
Dy      913 GCCTCTGTCGTTATTGGAAATTCTGGAAATGAGTGTTAGGTGCTCTTGCAAAA 972
Db      5966 ATACTGA - -- ATTGGATTCTAAAGAAATGAAATGAAATGAAATGAAATGAAATGAA 6021
Dy      973 AAAAAAAAGAAAAAAA 988
Db      6022 AATATAAACGGAAANAA 6037

RESULTS 21
AC103588          208670 bp DNA linear PRI 21-MAR-2002
LOCUS          Homo sapiens chromosome 3 clone RP11-640H5, complete sequence.
DEFINITION      AC103588
ACCESSION       AC103588.2 GI:119570161
KEYWORDS       HTG.
SOURCE         Homo sapiens (human)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 208670)
AUTHORS        Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
                Saenphimachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
TITLE          Unpublished
JOURNAL        2 (bases 1 to 208670)
AUTHORS        Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE          Direct Submission
JOURNAL        University of Washington,
DEFINITION      Submitted (29-NOV-2001) Genome Center, University of Washington,
ACCESSION       Box 352145, Seattle, WA 98195, USA
KEYWORDS       3 (bases 1 to 208670)
AUTHORS        Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
                Saenphimachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
TITLE          Direct Submission
JOURNAL        Box 352145, Seattle, WA 98195, USA
COMMENT        On Mar 21, 2002 this sequence replaced gi:17149452.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgctg@u.washington.edu
----- Project Information
Center project name: Chr-3
Center clone name: RP11-640H5 (bc0507)
----- Summary Statistics

```

----- 2292 2308 773 <800 2483 10346 10090 1901 1860
----- 2981 2978 1639 1735 88 <800 10998 10794 ----- 275 <800
----- 2132 2153 1523 1485 1731 1673 267 <800 2573 2610
----- 8772 8812 1514 1485 4039 3956 715 <800 6470 6445
----- 594 <800 62 <800 2376 2365 4942 4854 ----- 2385 2365
----- 4622 4505 1698 1735 96 <800 1034 1043 1119 1087
----- 7761 7875 1776 1735 2802 2806 3805 3782 3192 3074
----- 16778 17016 3527 3557 1989 1860 75558 7751 5056 5013
----- 1849 1834 7630 7751 651 <800 4678 4675 21 <800
----- 4534 4505 919 860 718 <800 1233 1213 4314 4314

Query Match Score 69.2; DB 9; Length 208670;
Best Local Similarity 63.3%; Pred. No. 5.7e-06;
Matches 124; Conservative 0; Mismatches 68; Indels 4; Gaps 1;

Qy 795 GAGGAGTTGGCGGGTCAAGGATGGGAAATGGGAGCTCAGAGGACGCCAAGGAGCC 854
Db 89973 GATTACTTTGGATGGGAGAATGGGAGATCTAAAGTCAGGATACTAAAGTAGCAT 90032
Qy 855 ATGTTAGAATGAACCGTCCAGAGAGCCAAAGCAGGCCATCAGGTGTC 914
Db 90033 ATTATGGATGACCACATCTAGAGTCATATGCAACATGCTTAATTAATTTGT 90092
Qy 915 ACTGTTGTTATGGATGGTCAATGCAAATGAGTGTTAGTGCTCTGCCAACAAA 974
Db 90093 ACT--GTATTGGATTATGCTAAAGAATAGATTTAGCTGCCTTGCCATAAAG 90148
Qy 975 AAAAAA..... 990
Db 90149 CARACAACAAAATA 90164

RESULT 22
AL353752 LOCUS AL353752
DEFINITION Human DNA sequence from clone RP11-497K15 on chromosome 9, complete sequence.
ACCESSION AL353752
VERSION AL353752.6 GI:13751335 HTG.
KEYWORDS SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 177921)
AUTHORS Tracey,A.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
COMMENT On Apr 21, 2001 this sequence version replaced gi:13445347
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sv:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPPB; Information on the WORMPPB database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormppb. This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/BGP/Ch9>.

RP11-497K15 is from the library RPCI-11-2 constructed by the group of Paster de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: PBACE3_6

This sequence is the entire insert of clone RP11-497K15. The true left end of clone RP11-5802 is at 128837 in this sequence. The true right end of clone RP11-5802 is at 80971 in this sequence.

FEATURES

source	/note="match: GSS: Em:AO131836"	/note="match: GSS: Em:AO131836"
	/repeat_region	/note="AluY5 repeat: matches 133. .307 of consensus"
		/note="AluY5 repeat: matches 1614. .12288
		/note="AluY5 repeat: matches 1646. .16717
		/note="MfTfJ repeat: matches 104. .187 of consensus"
		/note="LmB8 repeat: matches 16802. .17084
		/note="LmB8 repeat: matches 17504. .17660
		/note="Lm4 repeat: matches 17657. .18355
		/note="LmB3 repeat: matches 18357. .18656
		/note="Alu5 repeat: matches 19964. .20259
		/note="Alu5 repeat: matches 20604. .20912
		/note="L2 repeat: matches 20983. .21292
		/note="AluY repeat: matches 21321. .21428
		/note="FluA repeat: matches 22740. .22870
		/note="Mf repeat: matches 22890. .22940
		/note="L2 repeat: matches 22989. .23469
		/note="match: STS: Em:U67307"
		/repeat_region
		/note="match: STS: Em:U67307"
		/repeat_region
		/note="AluSg repeat: matches 24070. .24362
		/repeat_region
		/note="AluSg repeat: matches 24403. .25216
		/repeat_region
		/note="LmA6 repeat: matches 25217. .25355
		/repeat_region
		/note="LpB1 repeat: matches 25356. .25420
		/repeat_region
		/note="LmA6 repeat: matches 25415. .25744
		/repeat_region
		/note="LmA6 repeat: matches 26967. .27168
		/repeat_region
		/note="Mf repeat: matches 27190. .27327
		/repeat_region
		/note="Charlie1 repeat: matches 27323. .27475
		/repeat_region
		/note="Charlie1 repeat: matches 27607. .27917
		/repeat_region
		/note="AluS repeat: matches 28407. .29843
		/repeat_region
		/note="LmB7 repeat: matches 32140. .32448
		/repeat_region
		/note="AluSx repeat: matches 33717. .34022
		/repeat_region
		/note="AluS repeat: matches 34170. .34419
		/repeat_region
		/note="MfTg repeat: matches 34527. .34626
		/repeat_region
		/note="MfTf repeat: matches 35085. .35128
		/repeat_region
		/note="BpR1 repeat: matches 35525. .35836
		/repeat_region
		/note="AluS repeat: matches 36751. .37068
		/repeat_region
		/note="AluY repeat: matches 35844. .35911
		/repeat_region
		/note="MfTf repeat: matches 36507. .36760
		/repeat_region
		/note="LpB2 repeat: matches 36751. .37068
		/repeat_region
		/note="AluS repeat: matches 37069. .37794
		/repeat_region
		/note="MfTf repeat: matches 38795. .39098
		/repeat_region
		/note="AluJb repeat: matches 39059. .39146
		/repeat_region
		/note="LpB2 repeat: matches 39612. .39895
		/repeat_region
		/note="AluSx repeat: matches 1. .284 of consensus"
misc_feature	/note="match: GSS: Em:BS8243"	
misc_feature	/note="match: GSS: Em:AO515148"	
misc_feature	/note="match: GSS: Em:AO515148"	


```

repeat_region
repeat_region complement(70886..7283)
/rptc_family="MIR5A"
repeat_region /rptc_family="L12"
repeat_region complement(8426..8712)
/rptc_family="L1PA10"
repeat_region complement(8753..8917)
/rptc_family="MIR54A"
repeat_region complement(8984..9076)
/rptc_family="MIR"
repeat_region 9639..9748
/rptc_family="MIR"
repeat_region complement(9766..10614)
/rptc_family="MIR54A"
repeat_region 10700..10834
/rptc_family="L2"
repeat_region 10855..10967
/rptc_family="GA-rich"
repeat_region complement(11035..11307)
/rptc_family="L3"
repeat_region 11316..11349
/rptc_family="(TC)n"
repeat_region 11345..11391
/rptc_family="(CA)n"
repeat_region complement(11669..11975)
/rptc_family="MLT1G1"
repeat_region complement(12252..14205)
/rptc_family="MIR145"
repeat_region complement(14665..15444)
/rptc_family="MER114"
repeat_region 15495..15722
/rptc_family="L1ME1"
repeat_region 15815..16227
/rptc_family="L1P3"
repeat_region 16228..16256
/rptc_family="(TTTG)n"
repeat_region 16257..16678
/rptc_family="L1P3"
repeat_region 16679..17746
/rptc_family="MER11C"
repeat_region 17747..19574
/rptc_family="L1P3"
repeat_region 19575..21107
/rptc_family="L1PA7"
repeat_region complement(24451..24629)
/rptc_family="MIR"
repeat_region complement(24706..25065)
/rptc_family="L1PB1"
repeat_region 25066..25181
/rptc_family="(TA)n"
repeat_region 25182..25866
/rptc_family="L1"
repeat_region 25871..25976
/rptc_family="(TA)n"
repeat_region complement(25991..26108)
/rptc_family="L1P5"
repeat_region complement(26175..26288)
/rptc_family="L1PA5-16"
repeat_region complement(26305..26969)
/rptc_family="L1PA15-16"
repeat_region 26967..27157
/rptc_family="L1"
repeat_region complement(27164..27230)
/rptc_family="MIR61D"
repeat_region 27236..27400
/rptc_family="L1PA16"
repeat_region complement(27401..27697)
/rptc_family="AluSp"
repeat_region 27698..29368
/rptc_family="L1PA16"
repeat_region 29271..29293
/rptc_family="qual SNGL region"
/note=<20
repeat_region complement(29369..29564)

```

```

/reptc_family="MER61D"
29981..30691
/reptc_family="L1PA16"
31815..31838
/reptc_family="AT-rich"
32875..32910
/reptc_family="AT-rich"
32979..33163
/reptc_family="MLT1A1"
33164..33465
/reptc_family="AluJb"
33466..33616
/reptc_family="MLT1A1"
33704..34140
/reptc_family="MLT1K"
34145..34829
/reptc_family="MLT1K"
35312..35700
/reptc_family="L12"
36237..37252
/reptc_family="L1MC1"
complement(37557..38067)
/reptc_family="L1TR4a"
39342..39757
/reptc_family="MLT2A2"
39810..39818
/reptc_family="MLT2A2"
39819..39874
/reptc_family="MLT2A2"

Query Match 6.2% ; Score 61.8; DB 9;
Best Local Similarity 57.5%; Pred. No. 0.00035;
Matches 11; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 796 AGGAGTTGGGGGTCAAGGATAGGGAGGTCAAGGAGCCAAAGCAGGCCA 855
Db 100174 AGGAGCTGGAGGGCGGAGAATGGGGAGTAGGATGAGCTGAGATA 100115
Qy 856 TTGATGATGACCGTCAGGCAGCCAGCAAGCAAGCTGAGCTGATCAGGGCA 915
Db 100114 TGTAAATTAAACAATTCTGAGACCCATTGACCCATGAGATTAAGAT 100055
Qy 916 CTGATCTGATTGGAGTCATGCAAATGAGTCCTTGTAGCTGCTTCGCCACAAAAA 975
Db 100054 TGTACTGTTATTGGATTTGTCAATTGAGTATTGCTCTTGCCACARAC 99995
Qy 976 AAAAAAAAAAAA 988
Db 99994 AAAAAATTTAAAAAA 99982

RESULT 24
AF429315
LOCUS 125020 bp DNA linear PRI 18-JAN-2002
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION 1
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Homidae; Homo.
REFERENCE Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S., Ingersoll,Ashworth,R.G., Fleisher,A., Stevanini,G., Brice,A., Potter,N.T., Ross,C.A. and Margolis,R.L.
AUTHORS
TITLE A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.

```


COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 11, 2000 this sequence version replaced 91::7280308.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www.seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu

Center: Project name: L4198
 Center clone name: 276 N 1

----- Summary Statistics -----

Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 14809 bases at least Q40
 Consensus quality: 157187 bases at least Q30
 Consensus quality: 160494 bases at least Q20
 Insert size: 188000; agarose-EP
 Insert size: 162462; sum-of-contigs
 Quality coverage: 3.6 in Q20 bases; agarose-fP
 Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 24 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1210: contig of 1218 bp in length
 * 1219: gap of 100 bp
 * 1319: contig of 1065 bp in length
 * 2384: gap of 100 bp
 * 2484: contig of 1388 bp in length
 * 3872: gap of 100 bp
 * 3972: contig of 865 bp in length
 * 4837: gap of 100 bp
 * 4937: contig of 1325 bp in length
 * 6262: gap of 100 bp
 * 6362: contig of 1711 bp in length
 * 6333: gap of 100 bp
 * 8433: contig of 2938 bp in length
 * 11471: gap of 100 bp
 * 16137: contig of 4666 bp in length
 * 16237: gap of 100 bp
 * 20961: contig of 4724 bp in length
 * 21061: gap of 100 bp
 * 25287: contig of 4227 bp in length
 * 25387: gap of 100 bp
 * 32039: contig of 6652 bp in length
 * 32040: gap of 100 bp
 * 32140: contig of 6629 bp in length
 * 38769: gap of 100 bp
 * 38869: contig of 4916 bp in length
 * 43785: gap of 100 bp
 * 43884: contig of 6517 bp in length
 * 50401: gap of 100 bp
 * 50502: contig of 6581 bp in length
 * 57183: gap of 100 bp
 * 63917: contig of 6735 bp in length
 * 64017: gap of 100 bp
 * 71454: contig of 7437 bp in length
 * 71455: gap of 100 bp
 * 71555: contig of 8116 bp in length
 * 79671: gap of 100 bp
 * 79771: contig of 7643 bp in length
 * 87414: gap of 100 bp
 * 93773: contig of 6260 bp in length
 * 93873: gap of 100 bp

----- FEATURES -----

source 1. 164762
 /clone="RP11-276N1"
 /clone_lib="RCI-11 Human Male BAC"
 misc_feature 1..1218
 /note="assembly_fragment"
 misc_feature 1319..2383
 /note="assembly_fragment"
 misc_feature 2484..3671
 /note="assembly_fragment"
 misc_feature 3972..4336
 /note="assembly_fragment"
 misc_feature clone_end:SP6
 /vector_side:right"
 misc_feature 4937..6261
 /note="assembly_fragment"
 misc_feature 6362..8332
 /note="assembly_fragment"
 misc_feature 8433..11370
 /note="assembly_fragment"
 misc_feature 11471..16136
 /note="assembly_fragment"
 misc_feature 16237..20960
 /note="assembly_fragment"
 misc_feature 21061..25287
 /note="assembly_fragment"
 misc_feature 25388..32039
 /note="assembly_fragment"
 misc_feature 32140..38768
 /note="assembly_fragment"
 misc_feature 38869..43785
 /note="assembly_fragment"
 misc_feature 43885..50401
 /note="assembly_fragment"
 misc_feature 50502..57082
 /note="assembly_fragment"
 misc_feature 57183..63917
 /note="assembly_fragment"
 misc_feature 64018..71454
 /note="assembly_fragment"
 misc_feature 71555..79670
 /note="assembly_fragment"
 misc_feature 79771..87413
 /note="assembly_fragment"
 misc_feature 87514..93773
 /note="assembly_fragment"
 misc_feature 93874..103550
 /note="assembly_fragment"
 misc_feature 103691..113309
 /note="assembly_fragment"
 misc_feature 118410..113058
 /note="assembly_fragment"
 misc_feature 136159..164762
 /note="assembly_fragment"

----- BASE COUNT -----

ORIGIN 47298 a 33699 c 32793 g 48671 t 2301 others

Query Match 6.1%; Score 60.8%; DB 2; Length 164762;
 Best Local Similarity 64.1%; Pred. No. 0.0006;

Matches 109; Conservative 0; Mismatches 57; Indels 4; Gaps 1; Score 60.8; DB 9; Length 170898; Best Local Similarity 64.1%; Pred. No. 0.000; Mismatches 109; Conservative 0; Indels 4; Gaps 1;

Qy 814 GGATGGAAATGGGAGCTCAGGGCCAAAGGCGCTGTAGAATGAAACCTCCA 873
 Db 64641 GAAATGGGGAGTAGTGTAGCTCAGGGACAAAGTAACAGATGAGGATGACAGTCT 64700

Qy 874 GAGAGCCAGCAAGCAGGACTGAGGCTGAGGACTGTTGGAATTCGGATTGGAGTT 933
 Db 64701 AGAGATAAAATGAGCTAACGATAATTAATTTGACT---GTTATGGATI 64756

Qy 934 CATGCAAATATGAGTGTGTTAGCTGCCTGGCACAAAAAAA 983
 Db 64757 CCTGCTAAATGAGTAGACTTACGTTACGTTACAAACTACAAA 64806

RESULT 26
 AL353764/C LOCUS Human DNA sequence from clone RP11-440G5 on chromosome 9, complete sequence.
 DEFINITION PRI 01-JUN-2001
 AC104684
 ACCESSION AL353764
 VERSION GI:14272263
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 170898)
 AUTHORS Tracey,A.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail: enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 COMMENT On May 31, 2001 this sequence version replaced gi:13990016. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL: Sw:, SWISSPROT:, Tr:, TRAMBBL: Wp:, WORMPEP: Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> This sequence is from the library RP11-2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bapac/home.htm>

FEATURES source
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-440G5. The true left end of clone RP11-81B is at 105684 in this sequence. The true right end of clone RP11-44A5 is at 7076 in this sequence.
 Location/Qualifiers 1..170898

source
 /organism="Homo sapiens"
 /mol_type="Genomic DNA"
 /db_xref=taxon:9606"
 /chromosome="9"
 /clone="RP11-440G5"
 /clone_id="RP11-11_2"
 48782 a 34836 c 35124 g 52156 t

BASE COUNT
 ORIGIN

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, J., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 5:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.genegen.com>) or Pieter de Jong and coworkers at <http://www.chorix.org>

VECTOR: PBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is XXcos-2052C30; the clone sequenced to the right is RP1-356M6, 2000 bp overlap. Actual start of this clone is at base position 1 of RP1-1N7.

Sequence derived from one plasmid subclone, base position 94919 to 94957.

Data from AC067919 and AC097644 was used to finish this clone, AC104684.

Polymorphisms have been identified between AC067919, AC079779 and AC104684.

The sequence of AC021641 has been incorporated into AC104684.

FEATURES

Source

```
/rpt_family="AChobo"
3808_3939
/repeat_region
  /rpt_family="MIR"
  repeat_region
    /rpt_family="AT_rich"
    8128_8236
/repeat_region
  /rpt_family="CR1"
  repeat_region
    /rpt_family="MALR"
    10147_10658
/repeat_region
  /rpt_family="MALR"
  repeat_region
    11021_11493
/repeat_region
  /rpt_family="MALR"
  repeat_region
    11831_12002
/repeat_region
  /rpt_family="ERV1"
  repeat_region
    12654_12975
/repeat_region
  /rpt_family="ALu"
  repeat_region
    13127_13247
/repeat_region
  /rpt_family="MALR"
  repeat_region
    13553_13678
/repeat_region
  /rpt_family="ERV1"
  repeat_region
    13679_14042
/repeat_region
  /rpt_family="ERV1"
  repeat_region
    15284_15648
/repeat_region
  /rpt_family="MALR"
  repeat_region
    17182_17546
/repeat_region
  /rpt_family="MALR"
  repeat_region
    18734_18848
/repeat_region
  /rpt_family="MIR"
  repeat_region
    18831_19169
/repeat_region
  /rpt_family="ERV1"
  repeat_region
    19405_19541
/repeat_region
  /rpt_family="MIR"
  repeat_region
    19918_20175
/repeat_region
  /rpt_family="MALR"
  repeat_region
    21336_21561
/repeat_region
  /rpt_family="U1"
  repeat_region
    21567_22109
/repeat_region
  /rpt_family="ERV1"
  repeat_region
    23724_23845
/repeat_region
  /rpt_family="MIR"
  repeat_region
    23925_23987
/repeat_region
  /rpt_family="MIR"
  repeat_region
    24193_24232
/repeat_region
  /rpt_family="L2"
  repeat_region
    24355_24371
/repeat_region
  /rpt_family="MIR"
  repeat_region
    25391_26298
/repeat_region
  /rpt_family="ALu"
  repeat_region
    26273_26295
/repeat_region
  /rpt_family="AT_rich"
  repeat_region
    26380_26415
/repeat_region
  /rpt_family="L1"
  repeat_region
    26417_26752
/repeat_region
  /rpt_family="L1"
  repeat_region
    26754_27066
/repeat_region
  /rpt_family="L1"
  repeat_region
    27067_27345
/repeat_region
  /rpt_family="ALu"
  repeat_region
    27350_27667
/repeat_region
  /rpt_family="L1"
  repeat_region
    27457_27494
/repeat_region
  /rpt_family="(CA)n"
  repeat_region
    28171_28277
/repeat_region
  /rpt_family="(TA)n"
  repeat_region
    28336_28384
/repeat_region
  /rpt_family="AT_rich"
  repeat_region
    28376_28554
/repeat_region
  /rpt_family="L1"
  repeat_region
    28916_29170
/repeat_region
  /rpt_family="L1"
  repeat_region
    29509_29500
/repeat_region
  /rpt_family="L1"
  repeat_region
    29711_29749
/repeat_region
  /rpt_family="L1"
```

repeat_region 29751 . 30738

Query Match 6.1%; Score 60.4; DB 9; Length 129770;

Best Local Similarity 56.6%; Pried No. 0; 0.00076;

Matches 112; Conservative 0; Mismatches 86;

Indels 0; Gaps 0;

QY 794 AGGGAGGTTGGGGATTCTAGGATAGGGATGGGAGGTCAGGACAGCAGCC 853

DB 77435 AGGGCACTGTAGGGATGGGAGGATAGGGTAAAGGTAAAGTAGCAGG 77494

QY 854 CATTGAGAATGAACGCTCCAGGGACAGAACAGCGACAGACTCAGGCATAGCGTG 913

DB 77495 TGTGGGATAAACAGTCAGAGATCTCATGTACACATGAGCTAGTTAAAAAA 77554

QY 914 CACTGTCGATTGGATGGAGTCATGAAAATGAGTGTTAGCTGCTGCCACAAA 973

DB 77555 ATTGTGCTGTGATTGGATCTGTATAATGAGTTAGTTGCTATTACACAAA 77614

QY 974 AAAAAAAAAAAAAAAA 991

DB 77615 CARACACACACACAAA 77632

RESULT 28

AC067919/C AC067919 171655 bp DNA linear HTG_26-MAY-2000

DEFINITION Homo sapiens clone RP1-292I6, WORKING DRAFT SEQUENCE, 20 unordered pieces.

AC067919 AC067919_2 GI:8076830

VERSUS HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Primates; Catarrhini; Hominidae; Homo.

SOURCE Barren.B., Linton.L., Nusbaum,C. and Lander,E.

ORGANISM Homo sapiens (human)

JOURNAL 2 (bases 1 to 171655) published

REFERENCE 2 (bases 1 to 171655)

AUTHORS Barren.B., Linton.L., Nusbaum,C., Lander.E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna.N., Bastien,V., Beda,F., Beguslavsky,L., Boukhaite,B., Brown,A., Burkett,G., Campopiano,A., Castle.A., Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cooke,P., DeReilano,K., Devar.K., Diaz,J.S., Dodge,S., Domino.M., Doyle.M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Giande,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine.R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marcus,N., McCarthy,M., McEwan,P., McCorkell,A., McPheeeters,R.,,, Medrano.J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Murphy,T., Maylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanu,C., Polaris,Z., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talama,J., Testayre,S., Theodore,J., Turrell,A., Travers,M., Trigilio,J., Vassilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

JOURNAL Direct Submission (22-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On May 25, 2000 this sequence version replaced gi:762358. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE JOURNAL

FEATURES misc_feature

SOURCE

Location/Qualifiers

1. .171655 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone_id="RP1-292I6" /clone_end="RP1-292I6" /note="assembly_fragment clone_end:T7 vector_side:left"

misc_feature	321..4001	ORGANISM	Homo sapiens
	/note="assembly_fragment"		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	4102..7945		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
misc_feature	8046..11450	REFERENCE	1 (bases 1 to 124347)
	/note="assembly_fragment"	AUTHORS	Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T., and Hood,L.
misc_feature	11551..16117	TITLE	Sequencing of human chromosome 14
	/note="assembly_fragment"	JOURNAL	Unpublished
misc_feature	16218..20480	REFERENCE	2 (bases 1 to 124347)
	/note="assembly_fragment"	AUTHORS	Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T., and Hood,L.
misc_feature	20581..25760	TITLE	Submitted (11-SEP-1999) Multimegabase Sequencing Center, University
	/note="assembly_fragment"	JOURNAL	of Washington, PO BOX 357730, Seattle, WA 98195, USA
misc_feature	25861..31039	REFERENCE	3 (bases 1 to 124347)
	/note="assembly_fragment"	AUTHORS	Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T., and Hood,L.
misc_feature	31140..35620	TITLE	Direct Submission
	/note="assembly_fragment"	JOURNAL	Submitted (19-NOV-1999) Multimegabase Sequencing Center, University
misc_feature	35921..42490	COMMENT	of Washington, PO BOX 357730, Seattle, WA 98195, USA
	/note="assembly_fragment"	ON	On Nov 19, 1999 this sequence replaced gi:6114900.
misc_feature	42591..49780	COMMENT	-----
	/note="assembly_fragment"	Center	Genome Center
misc_feature	4981..57796	CODE	UMSC
	/note="assembly_fragment"	WEB SITE	http://chroma.mbt.washington.edu/msg_www
misc_feature	57897..67541	CONTACT	leezrowen@washington.edu
	/note="assembly_fragment"	-----	Summary Statistics
misc_feature	67642..76045	SEQUENCING VECTOR	PUC18; L08752
	/note="assembly_fragment"	CHEMISTRY	Big Dye Terminators and primers
misc_feature	76146..83834	ASSEMBLY PROGRAM	Phrap; version 0.990399
misc_feature	85935..95701	FEATURES	-----
misc_feature	95802..107607	SOURCE	Location/Qualifiers
	/note="assembly_fragment"	1..124347	
misc_feature	107708..123352	/clone	/clone=CTD-217314
	/note="assembly_fragment"	LIBRARY	/clone.lib="CalTech human BAC library P11-114N19,"
misc_feature	123453..141544	OVERLAPS	/note="This clone overlaps RP11-114N19 , Accession AC007262"
misc_feature	141645..171655	CDS	complement(join(42575..42673,43714..43786,47934..48025,52809..52927,53951..54077,62318..62371))
ORIGIN	50815 a 35321 c 34768 g 48847 t 1904 others		/note="Intron-exon boundaries defined in relation to EST
BASE COUNT			AL554064.. the closest BLASTX similarity is to a drosophila potassium channel protein
			/codon_start=1
			/product="unknown"
			/protein_id="PAAFO9033..1"
			/db_xref="GI:6453845"
			/transl_table=10
			/translation="MLGRYKEISNGQAEIHLKESUDIDLRSORLIRNSGGRSI
			GFIHQSLRDLSEISQRLGDFNFEELRSRSRDATEKRALEELTERKLN
			EQKVKIQCLSSAFPDVNVPINWQCGI"
			958810..958860
			/note="low quality data"
			103634..>103846
			/product="thyroid stimulating hormone receptor"
			/note="This is the 5' end of TSHR, found in M73745."
RESULT 29			103677..>103846
AC010072	124347 bp	DEFINITION	/product="TSHR"
LOCUS	DNA linear	AC010072	/db_xref="GT:6453844"
DEFINITION	14q11 clone CTD-217314 containing TSHR	VERSION	/translation="MRPADLQLQVLLDPLGGMCSSPPCEHQEDDFRVTCKDI
ACCESSION	AC010072	KEYWORD	ORIPSLPPSTQNL"
VERSION	GI:6453843	SOURCE	
KEYWORD	HTG,		
KEYWORD	Homo sapiens (human)		

misc_feature
 109619..124347
 /note="Overlap with RP11-114N19, Accession AC007262"
 variation
 110216
 /note="217314: c; 114N19: t"
 /replace="t"
 variation
 110755
 /note="217314: g; 114N19: a"
 /replace="a"
 variation
 111115
 /note="217314: c; 114N19: g"
 /replace="g"
 variation
 114198
 /note="217314: g; 114N19: a"
 /replace="a"
 variation
 114590
 /note="217314: c; 114N19: t"
 /replace="t"
 variation
 115183
 /note="217314: t; 114N19: tata"
 /replace="taat"
 115110..115200
 /note="Low quality data"
 variation
 116847
 /note="217314: t; 114N19: c"
 /replace="c"
 variation
 117131
 /note="217314: g; 114N19: a"
 /replace="a"
 variation
 117408
 /note="217314: a; 114N19: c"
 /replace="c"
 variation
 117633
 /note="217314: g; 114N19: t"
 /replace="t"
 variation
 118901
 /note="217314: a; 114N19: g"
 /replace="g"
 variation
 118988
 /note="217314: c; 114N19: t"
 /replace="t"
 variation
 119250
 /note="217314: a; 114N19: g"
 /replace="g"
 variation
 119718
 /note="217314: g; 114N19: t"
 /replace="t"
 variation
 120914
 /note="217314: c; 114N19: a"
 /replace="a"
 variation
 121395..121413
 /note="217314: caaaaaaaaaaaaa; 114N19: c"
 /replace="c"
 variation
 122019
 /note="217314: g; 114N19: a"
 /replace="a"
 variation
 123721
 /note="217314: c; 114N19: t"
 /replace="t"
 variation
 123900
 /note="217314: t; 114N19: c"
 /replace="c"
 variation
 123926..123927
 /note="217314: tt; 114N19: t"
 /replace="t"
 variation
 124015
 /note="217314: t; 114N19: c"
 /replace="c"
 variation
 124182
 /note="217314: g; 114N19: a"
 /replace="a"

Query Match 5.9%; Score 58.8; DB 9; Length 124347;
 Best Local Similarity 61.8%; Pred. No. 0.0018; Mismatches 67; Indels 1; Gaps 1;
 Matches 110; Conservative 0; MisMatches 67; Indels 1; Gaps 1;

Qy 814 GGATAGGGATGGGAGGTAGAGGAAGGAGCCATGATGAACTGACCGTCCA 873
 Db 41340 GAAATGGGAGATGGATGAGGATCAAAAGTAGCTATGATGAACTGATCTAGGGTAAAGTC 41399

Qy 874 GAGGCCAAGCACGCCAGGGACTCAGGGCCATAGGGTGCAGTGCTATTGGAGTT 933
 Db 41400 AGGATCTAAAGTATACTGAGACTACAGATAATAAAACTGTACT ATATATGGGATT 41458

Qy 934 CATGAAATGAGGTGTTAGCTGCTTGTGCTTGTGTTCTGGCACACCAAC 991
 Db 41459 CATTGCTAGTAACTACATTAGCTGTTCTGGCACACCAAC 41516

RESULT 30
 AC083574/C AC093574 151993 bp DNA linear PRI 23-FBB-2002

LOCUS Homo sapiens chromosome 3 clone RP11-178K11, complete sequence.

DEFINITION AC093574 AC021597

ACCESSION AC093574.2 GI:18873861

VERSION HTIG.

KEYWORDS Homo sapiens (human)

ORGANISM Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 151993)

REFERENCE Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and Haugen,E.D.

AUTHORS

VERSION

TITLE Unpublished

JOURNAL 2 (bases 1 to 151993)

REFERENCE Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and Haugen,E.D.

AUTHORS

JOURNAL Direct Submission

COMMENT Submitted (05-SEP-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

REFERENCE 3 (bases 1 to 151993)

AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.

JOURNAL Direct Submission

COMMENT Submitted (23-FEB-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

On Feb 23, 2002 this sequence version replaced gi:15431100.

COMMENT ----- Genome Center

Center: University of Washington Genome Center

Center Code: UWGC

Web site: <http://www.genome.washington.edu>

Drafting Center: BCM

Contact: wgctgs@bcm.edu

----- Project Information

Center project name: chr-3

Center clone name: RP11-178K11 (bc0286)

----- Summary Statistics

Sequencing vector: Unknown; 56% of reads

Sequencing vector: Plasmid; 44% of reads

Chemistry: Dye-Terminator ET; 49% of reads

Chemistry: Dye-Terminator Big Dye; 51% of reads

Assembly program: Phrap; version 0.90319

Consensus quality: 151851 bases at least Q40

Consensus quality: 151973 bases at least Q30

Consensus quality: 151993 bases at least Q20

Insert size: 156451; sum-of-contigs

Quality coverage: 8.9x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5': RP11-56B20 (JWGC:bc0174) AC026673

3': Mapping in progress

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality

Query Match Similarity 61.0% ;保守性 Matches 114; Score 5.9%; Best Local Similarity 61.0% ; Pred. No. 0.0021; Mismatches 0; Gaps 1;	COMMENT
Qy 797 GSAGTTGGGGTTCAGGATAGGGATGGGAGGTCAAGGACGCCAAAGCAGCCAT 856	
Db 36052 GGATGGGGAGGGCAGGAAATGAGGAATGAGTGAAGAATAAAAGGGCAAT 35993	
Qy 857 GTAGAAATGAAACCGTCCAGAGGAGCCAAAGCAGGGCTGAGGACTTGAGCTGCAC 916	
Db 35992 GTATGATTAAACAGGTCAAGTCAATAACATGAAATTAGCTAAATGTGTC 35933	
Qy 917 TGTTCGTTATGGAGCTCATGCAAATGAGTGTTTAGTGCTCTGGCCACAAAAAA 976	
Db 35932 T---GTATGAGATTCTGTTAACATGAGATTTACCTGCTTGGCACGAAACX 35877	
Qy 977 AAAAAA 983	
Db 35876 AACAGAA 35870	
RESULT 31 AC012087	LOCUS
DEFINITION Homo sapiens 3 BAC RP11-311L10,RP11-150K6 RP11-68D12 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	DNA
ACCESSION AC012087 AC011608 AC012034 AC012016	PRIM.
KEYWORDS FRTG.	
SOURCE Homo sapiens (human)	ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	AUTHORS
Muzny D.M., Adams C.M., Bailey,M., Barbaria,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buahy,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domani-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Gorrell,L.L., Guevara,W., Harris,K., He,X., Hernandez,U., Hodgson,A., Hogues,M., Holloway,C., Hobak,H., Gorrell,J.H., Jackson,L., Jia,Y., Jones,M., Kondrejkowski,J., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarzge,O., Liu,J., Liu,W., Logan,O., Lczado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Moore,S., Moorish,T., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.I., Quiles,M., Reiter,D., Rivers,M., Samuel,S., Say,J., Shab,E., Scheerer,S., Shen,H., Simon,M., Sparks,A., Stamts,A., Sugang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahban,M., Washington,S., Wrenstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrenstock,G., Yu,W., Zhou,X., Naylor,S.L., Nelson,D. and Gibbs,R.	
TITLE	JOURNAL
REFERENCE	AUTHORS
Worley,K.C.	TITLE
Direct Submission	JOURNAL
Submitted (20-Oct-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
REFERENCE	AUTHORS
Worley,K.C.	TITLE
Direct Submission	JOURNAL
Submitted (01-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
REFERENCE	AUTHORS
Worley,K.C.	TITLE
Direct Submission	JOURNAL
Submitted (18-Aug-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One	

Baylor Plaza, Houston, TX 77030, USA
On or before Aug 18, 2000 this sequence version replaced
91:16067112, gi:7025646, gi:7025648.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:1389-1402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://grc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

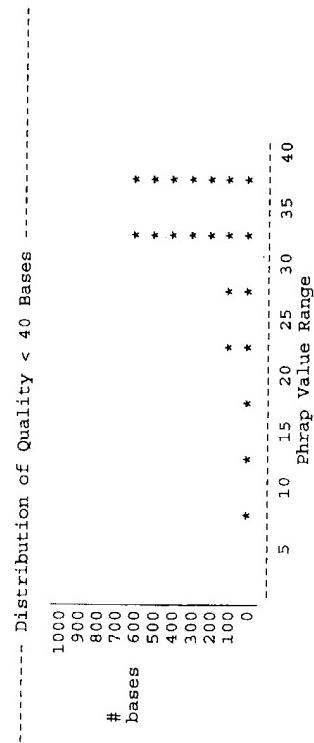
----- Summary Statistics -----

Contig length: 194635 Phrap values in estimate: 194635
Phrap values in estimate: 193865
Average error rate (BCM-Phrap estimate): 3.31225e-05
Fraction of Phrap values less than 40 : 0.00865035
Number of consensus changing edits: 46
Number of N's in consensus : 0

----- Consensus changing edits -----

Position Original+Context Edited+Context

4981 agatggat(n) tgatggatgg agaaacttta(a) ccaccaggcc tctggggcc(c) taatggatgt ctgaatttgaa(n) ttccatataaa aaaaaaaaaa(a) tcattggccaa tgcaaaaaag(t) agatggatgc ttcaatggtg(g) tattttgtgtt caggccccc(c) ccccccgtt ccccccgtt(c) gggccccggg tgctctaa(g) tcctgtggcc aagaadagg(g) (a) gggtggggcc gcttgtgcct(n) tcgttgatga caatctgtg(t) agatgtgtt aagcgtatgt(a) ctcatcttt ttttttttttttttttttt ataaaggatgg(c) ctttgtgtttt gtaaaaatttg(g) taaaatggaaat ggatggctg(c) aaaaatggcc cgtatggccgg(n) gggatggcc



repeat_region /note="L1MC3 repeat: matches 6433. . 6606 of consensus" 1745. . 2410
repeat_region /note="L1MC3 repeat: matches 6609. . 7281 of consensus" 2411. . 2705
repeat_region /note="Alusq repeat: matches 1. . 296 of consensus" 2706.
repeat_region /note="L1MC3 repeat: matches 7281. . 7736 of consensus" 3183. . 3479
repeat_region /note="Alusq repeat: matches 1. . 294 of consensus" complement(4350 4820"
repeat_region /note="match: GSS: Em:AQ827082" 4999. . 5225
repeat_region /note="L12 repeat: matches 1328. . 1564 of consensus" 5811. . 6105
repeat_region /note="Alusx repeat: matches 1. . 295 of consensus" 6180. . 6371
repeat_region /note="L1MC3 repeat: matches 5725. . 5926 of consensus" 6386. . 6421
repeat_region /note="L18 copies 2 mer gt 97% conserved" 6673. . 7475
repeat_region /note="L12 repeat: matches 1777. . 2614 of consensus" 7485. . 7709
repeat_region /note="MIR repeat: matches 21. . 262 of consensus" 7733. . 7783
repeat_region /note="L12 repeat: matches 2648. . 2705 of consensus" 8206. . 8517
repeat_region /note="Alusx repeat: matches 1. . 312 of consensus" 8833. . 10173
repeat_region /note="ptrs5 repeat: matches 743. . 2438 of consensus" 10349. . 10386
repeat_region /note="L19 copies 2 mer tg 100% conserved" 10388. . 10664
repeat_region /note="Aluy repeat: matches 2. . 279 of consensus" 11826. . 12245
repeat_region /note="Alusx repeat: matches 1. . 294 of consensus" 12656. . 12755
repeat_region /note="MIR repeat: matches 50. . 157 of consensus" 12776. . 13081
repeat_region /note="Alusq repeat: matches 1. . 306 of consensus" complement(13145. . 13540)
repeat_region /note="Alujo repeat: matches 1. . 258 of consensus" 13457. . 13719
repeat_region /note="Alujo repeat: matches 1. . 258 of consensus" 14443. . 14562
repeat_region /note="L12 copies 20 mer 78% conserved" 14515. . 14558
repeat_region /note="L22 copies 2 mer ag 773. conserved" complement(join(<14896. . 15073,19597. . 19738,40166. . 40316,
41154. . 41246,57685. . 57781,58449. . 59058,78822. . 78933,
80037. . 80147,88401. . 88572,97782. . 97814,97914. . 98057,
107109. . 107254))
repeat_region /note="match: CDNAs: Em:AF096286 Em:AB018348
Tr:Q9W506 Tr:O61825 Sw:PIB490"
repeat_region /codon_start=1
repeat_region /evidence="not experimental"
repeat_region /product="dJ1016N21_1"
repeat_region /evidence="Drosophila Pecanex (PCX), similar to KIAA0805)"
repeat_region /protein_id="CA2D20144_1"
repeat_region /db_xref="IGI:18070856"
repeat_region /translation="MVSQDLEASSCSSTQKFNREQFYKEIIFPGKWKWYDRDLIL
OPDPASPIKHQNQIQLITYSPYFCVTCGHLTSGQPKDMVYLFLCLMSQYSLIKSV
QSDLPASDIDENKVALLIVSLLGFLTSQGPKDMVYLFLCLMSQYSLIKSV
RSVLAALLGHAVCFSAVKEMWSMOMIPALPSAFCCLLVSYHLSROSSDPSVMSFI
QCRLPFKEKHONLAAADAELPKKOMDSUTDVLMNDLIVCAYVHLSFVASVATVFL
LRPFLSIVPALAGAVGFVTHVPLQRKHPWWMISHPLKRNKEYHOREVRDYLHM

CDS gene /evidence-not experimental
repeat_region /product="dJ1016N21_1 (novel protein (ortholog of Drosophila Pecanex (PCX), similar to KIAA0805))"
repeat_region /protein_id="CA2D20144_1"
repeat_region /db_xref="IGI:18070856"
repeat_region /translation="MVSQDLEASSCSSTQKFNREQFYKEIIFPGKWKWYDRDLIL
OPDPASPIKHQNQIQLITYSPYFCVTCGHLTSGQPKDMVYLFLCLMSQYSLIKSV
QSDLPASDIDENKVALLIVSLLGFLTSQGPKDMVYLFLCLMSQYSLIKSV
RSVLAALLGHAVCFSAVKEMWSMOMIPALPSAFCCLLVSYHLSROSSDPSVMSFI
QCRLPFKEKHONLAAADAELPKKOMDSUTDVLMNDLIVCAYVHLSFVASVATVFL
LRPFLSIVPALAGAVGFVTHVPLQRKHPWWMISHPLKRNKEYHOREVRDYLHM

Query Match 5.8%; Score 58; DB 9;
Best Local Similarity 59.7%; Pred. No. 0.0029;
Matches 117; Conservative 0; Mismatches 75; Indels 4; Gaps 1;

Oy 789 AGTGAAGAGGAGGTGTTGGGGTTCTAGGATAGGGATGGGAGGTAGAGGACGAAAGCA 84.8
Db 47275 AGTAACTATGGCATGGGGAGGTGAGATGGGGATATGTTGTTAGAATTATAAGTA 47324

Oy 849 GCAGGCCTGTAGAACCTGGTCCAGAGGAGCCAAGCACGGCAGGGCATCA 908
Db 47335 GCAGATATGTAGATAAACACTCCAGGATCTATGTCATGAGAACATAATTAA 47394

Qy 909 GCGTGCAGCTGATTTGGAGTCATGCCAAATGAGTGTCTTGCCA 968
 Db 47395 AACTGTACT---ATAATTGGATTCCCTGCCAATAAGTAGTTGGCCA 47450

Qy 969 CAAAAAAAATAAAAA 984
 Db 47451 CAAACACATATAAA 47466

RESULT 3.6

AC019159/c AC019159 163085 bp DNA linear PRI 07-NOV-2001
 DEFINITION Homo sapiens BAC clone RP11-56018 from 2, complete sequence.
 ACCESSION AC019159
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE AUTHORS Sulston,J.E. and Waterston,R.
 TITLE Toward a complete human genome sequence
 JOURNAL GENOME Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE AUTHORS Goyea,E., Cotton,M., Spalding,L. and Lehnert,L.
 TITLE The sequence of Homo sapiens BAC clone RP11-56018
 JOURNAL Unpublished

REFERENCE 3 (bases 1 to 163085)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 163085)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 5 (bases 1 to 163085)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-APR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 6 (bases 1 to 163085)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Apr 19, 2001 this sequence version replaced gi:11276269.
 --- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@genotson.wustl.edu
 --- Summary Statistics
 Center Project name: H_NH0056018

from more than one subclone, and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is AC027111; the clone sequenced to the right is AC023003. Actual start of this clone is at base position 1 of RP11-56018; actual end is at base position 163085 of RP11-56018.

The sequence from base position 822245 to 82358 is derived from a single plasmid subclone. Size information based on PCR supports the assembly.

FEATURES

source	Location/Qualifiers
	1. 163085
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="Taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-56018"
	/clone_lib="RPCI-11"
	1. .51
	/note="similar to EST AV696391 (NID:gi0298254)"
	107. .24
	/note="similar to EST AV696391 (NID:gi0298254)"
	197. .373
	/rpt_family="MIR"
	568. .667
	/rpt_family="MIR"
	668. .775
	/rpt_family="MER1_type"
	1249. .1439
	/rpt_family="MIR"
	8077. .8445
	/rpt_family="MIR"
	8451. .9804
	/rpt_family="MIR"
	9861. .10142
	/rpt_family="MIR"
	10149. .10487
	/rpt_family="MIR"
	10395. .10790
	/rpt_family="MIR"
	11489. .11710
	/rpt_family="MIR"
	11985. .12022
	/rpt_family="TG"
	12131. .12886
	/rpt_family="L1"
	13153. .13392
	/rpt_family="L1"
	13397. .13441
	/rpt_family="TG"
	13513. .13577
	/rpt_family="TA"
	13578. .14029
	/rpt_family="L1"

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence

repeat_region	14082 .14306 /rpt family="L1"	Oy 847 CAGCAGCCGATGAACTGAGAACC-GTCGAGAGGCCAAGCCGGCAAGGACTGGCAGCCCA 905
repeat_region	14739 .14778 /rpt family="L2"	Db 76473 CAGCAGCCGATGAGGAGGACCAAGTCAAGGATCAATTAGGAGGACCATACAG 76414
repeat_region	15455 .15476 /rpt family="AT_rich"	Oy 906 TCAAGCGTGCACTGTTCATTGGAGTCATGCAAATGAGTGTGTTAGTGCTTGTGCTTGTG 965
repeat_region	16189 .16500 /rpt family="Alu"	Db 76413 GTAAATAAGGGCTACTGTGAAGGANTCAGGCTAACTATA-GATTTGGTGCCTTGTG 76355
repeat_region	16601 .16736 /rpt family="MIR"	Oy 966 CCACAAAPAAAAAAAPAAAAA 987
repeat_region	19100 .1909 /rpt family="MIR"	Db 76354 CCACAAAPACACACAAACAAA 76333
repeat_region	19257 .19388 /rpt family="MIR"	
repeat_region	19840 .19806 /rpt family="TGAA)n"	
repeat_region	20239 .20271 /rpt family="(CA)n"	
repeat_region	20643 .20923 /rpt family="Alu"	
repeat_region	21036 .21437 /rpt family="MAlR"	
repeat_region	21463 .21653 /rpt family="MIR"	
repeat_region	21736 .21861 /rpt family="MIR"	
repeat_region	22171 .22330 /rpt family="Alu"	
repeat_region	22695 .22809 /rpt family="MIR"	
repeat_region	23975 .24015 /rpt family="A-rich"	
repeat_region	24958 .25194 /rpt family="MIR"	
repeat_region	25352 .25425 /rpt family="L1"	
repeat_region	26112 .26117 /rpt family="L1"	
repeat_region	26739 .2649 /rpt family="TA)n"	
repeat_region	28668 .28791 /rpt family="TA)n"	
repeat_region	29118 .29628 /rpt family="L2"	
repeat_region	30306 .30338 /rpt family="TA)n"	
repeat_region	30339 .30881 /rpt family="L1"	
repeat_region	30882 .30904 /rpt family="TA)n"	
repeat_region	30965 .31365 /rpt family="Alu"	
repeat_region	31375 .31426 /rpt family="MIR"	
repeat_region	31541 .31626 /rpt family="Alu"	
repeat_region	32257 .32329 /rpt family="MIR"	
repeat_region	32646 .33176 /rpt family="L2"	
repeat_region	33297 .33445 /rpt family="L2"	
repeat_region	33566 .33749 /rpt family="MIR"	
repeat_region	34342 .34453 /rpt family="MIR"	
Query Match	5.8%; Score 58; DB 9; Length 163085;	
Best Local Similarity	61.9%; Pred. No. 0.0029; Mismatches 75; Indels 2; Gaps 2;	
Matches	125; Conservative 0; Mismatches 75; Indels 2; Gaps 2;	
Qy	787 TGAATGAGGGAGTTGGGGGTTCAGGAGGTCAGGACGAAAG 846	
Db	76533 TGCAGAGGGTAGAAATGATAGGAATAGGACACAAAG 76474	

FEATURES source

Matches 112; Conservative 0; Mismatches 47; Indels 6; Gaps 2; complement(1507 .1926)
 /note="match: GSS: Em: AQ146487"
 misc_feature
 1822 .2311
 /note="match: GSS: Em: AQ624491"
 repeat_region
 1830 .1956
 /note="L2 repeat: matches 2616 .2746 of consensus"
 repeat_region
 2127 .2181
 /note="L2 repeat: matches 2656 .2710 of consensus"
 repeat_region
 2380 .2691
 /note="AluX repeat: matches 1 .311 of consensus"
 repeat_region
 2716 .3083
 /note="MUTIR repeat: matches 1 .390 of consensus"
 misc_feature
 complement(2776 .3233)
 /note="match: SNS: Em: G37725
 match: GSS: Em: B30422"
 misc_feature
 3252 .3690
 /note="match: GSS: Em: AQ773761"
 repeat_region
 4104 .4288
 /note="MIR repeat: matches 8 .219 of consensus"
 repeat_region
 4533 .4756
 /note="L2 repeat: matches 1992 .2228 of consensus"
 repeat_region
 4915 .5144
 /note="MIR repeat: matches 2 .252 of consensus"
 repeat_region
 5284 .5483
 /note="L2 repeat: matches 2541 .2749 of consensus"
 repeat_region
 7056 .7103
 /note="12 copies 4 mer tcca 85% conserved"
 repeat_region
 7216 .7391
 /note="MIR repeat: matches 73 .262 of consensus"
 repeat_region
 8076 .8276
 /note="MIR repeat: matches 61 .262 of consensus"
 repeat_region
 8458 .8494
 /note="MIR repeat: matches 31 .67 of consensus"
 repeat_region
 13489 .13811
 /note="L2 repeat: matches 1301 .1648 of consensus"
 repeat_region
 14392 .15214
 /note="MER45B repeat: matches 1 .835 of consensus"
 repeat_region
 16024 .1624
 /note="MIR repeat: matches 34 .248 of consensus"
 repeat_region
 16428 .16737
 /note="AluY repeat: matches 1 .299 of consensus"
 misc_feature
 complement(17098 .17761)
 /note="match: GSS: Em: AQ472707"
 repeat_region
 17714 .18068
 /note="L2 repeat: matches 2155 .2537 of consensus"
 repeat_region
 18073 .18710
 /note="L2 repeat: matches 1049 .1745 of consensus"
 misc_feature
 complement(18332 .19056)
 /note="match: GSS: Em: AQ27362"
 misc_feature
 complement(18876 .19056)
 /note="match: GSS: Em: AQ09729"
 misc_feature
 complement(18876 .19056)
 /note="match: STS: Em: HSPE58B10"
 misc_feature
 1880 .19060
 /note="match: STS: Em: HSPE23B05"
 repeat_region
 20092 .20163
 /note="L2 repeat: matches 2673 .2746 of consensus"
 repeat_region
 20198 .20725
 /note="LTRB repeat: matches 189 .691 of consensus"
 repeat_region
 20726 .20992
 /note="AluX repeat: matches 1 .299 of consensus"
 repeat_region
 20993 .21176
 /note="LTRB repeat: matches 1 .189 of consensus"
 repeat_region
 21234 .21886
 /note="LTR1 repeat: matches 1 .660 of consensus"
 repeat_region
 22078 .22430
 /note="LTR1C repeat: matches 3 .385 of consensus"
 repeat_region
 22651 .23063
 /note="MLTIC repeat: matches 50 .466 of consensus"
 repeat_region
 23449 .23783
 /note="MER2 repeat: matches 1 .345 of consensus"

FEATURES location

Location/Qualifiers

1 .119653
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q24 .3-25 .3"
 /clone="RP1-44G5"
 /clone_lib="RPCI-1"
 1 .237
 /note="AluSx repeat: matches 1 .237 of consensus"
 240 .925
 /note="L2 repeat: matches 1995 .2706 of consensus"

misc_feature	complement(24518..-25322)		Best Local Similarity 56.6%; Pred. No. 0.0032;
repeat_region	/note="match: GSS: Em:RQ745887"	/note="match: GSS: Em:RQ745887"	0; Mismatches 82; Indels 0; Gaps 0;
repeat_region	24705..24962	24705..24962	
repeat_region	/note="MIR repeat: matches 49..262 of consensus"		
repeat_region	30110..30149	30110..30149	
repeat_region	/note="20 copies 2 mer 99 77% conserved"		
repeat_region	30411..30718	30411..30718	
misc_feature	/note="AluY repeat: matches 1..299 of consensus"		
repeat_region	30754..31346	30754..31346	
repeat_region	/note="match: GSS: Em:RQ392308"		
repeat_region	30834..31024	30834..31024	
repeat_region	/note="MER58A repeat: matches 26..207 of consensus"		
repeat_region	31161..31780	31161..31780	
repeat_region	/note="MIR repeat: matches 2..115 of consensus"		
repeat_region	31784..32115	31784..32115	
repeat_region	/note="LIMA6 repeat: matches 5960..6294 of consensus"		
repeat_region	32217..32610	32217..32610	
repeat_region	/note="MIR repeat: matches 100..194 of consensus"		
repeat_region	32626..32939	32626..32939	
repeat_region	/note="Alusp repeat: matches 1..313 of consensus"		
repeat_region	33215..33494	33215..33494	
repeat_region	/note="LIPAS repeat: matches 5922..6145 of consensus"		RESULT 39
repeat_region	33620..33844	33620..33844	
repeat_region	/note="MIR repeat: matches 20..261 of consensus"		
repeat_region	33976..34701	33976..34701	
repeat_region	/note="LIR8 repeat: matches 1..691 of consensus"		
repeat_region	35234..35696	35234..35696	
repeat_region	/note="MLTIC repeat: matches 1..466 of consensus"		
repeat_region	35981..36020	35981..36020	
repeat_region	/note="10 copies 4 mer gaga 92% conserved"		
repeat_region	36393..36488	36393..36488	
repeat_region	/note="L2 repeat: matches 2682..2746 of consensus"		
repeat_region	36546..36711	36546..36711	
repeat_region	/note="MER5B repeat: matches 2..173 of consensus"		
repeat_region	37578..37810	37578..37810	
repeat_region	/note="AluSGJx repeat: matches 77..299 of consensus"		
repeat_region	37844..37945	37844..37945	
repeat_region	/note="L12 repeat: matches 2589..2710 of consensus"		
misc_feature	37938..38503	37938..38503	
misc_feature	/note="match: GSS: Em:RQ77965"		
repeat_region	38503..38560	38503..38560	
repeat_region	/note="match: GSS: Em:RQ721135"		
repeat_region	3920..3939	3920..3939	
repeat_region	/note="MIR repeat: matches 90..220 of consensus"		
repeat_region	40557..41199	40557..41199	
repeat_region	/note="L2 repeat: matches 1882..2529 of consensus"		
repeat_region	41533..42009	41533..42009	
repeat_region	/note="match: GSS: Em:RQ403526"		
repeat_region	42947..43072	42947..43072	
repeat_region	/note="MER5A repeat: matches 64..189 of consensus"		
repeat_region	43714..43818	43714..43818	
repeat_region	/note="L2 repeat: matches 2374..2478 of consensus"		
repeat_region	44053..44085	44053..44085	
repeat_region	/note="L2 repeat: matches 2673..2705 of consensus"		
repeat_region	44219..44274	44219..44274	
repeat_region	/note="28 copies 2 mer tt 78% conserved"		
repeat_region	44275..44669	44275..44669	
repeat_region	/note="Alusp/q repeat: matches 181..243 of consensus"		
repeat_region	45141..45400	45141..45400	
repeat_region	/note="Alusc repeat: matches 1..265 of consensus"		
repeat_region	45402..45515	45402..45515	
repeat_region	/note="MER5A repeat: matches 51..187 of consensus"		
repeat_region	46114..46312	46114..46312	
repeat_region	/note="L1M52 repeat: matches 5962..6154 of consensus"		
repeat_region	46321..47484	46321..47484	
repeat_region	/note="LIME repeat: matches 4307..5529 of consensus"		
repeat_region	47489..48076	47489..48076	
Query Match	5.8%; Score 57.8; DB 9; Length 119853;	FEATURES source	

Location/Qualifiers
1..138370

FEATURES source
1..138370

Query Match

5.8%; Score 57.8; DB 9; Length 119853;

COMMENT

JOURNAL

VERSION

TITLE

ORGANISM

REFERENCE

AUTHORS

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

KEYWORDS

SOURCE

PICTURE

IMAGE

COMMENT

DEFINITION

ACCESSION

LOCUS

DEFINITION

SEQUENCE

VERSION

887 GCGAGGACTCAGGCCATGGTGCACTTCTGATTGGAGTCATGAAATGAG 946
854 ACATGGAGCTTAGTAATAATGTACT---GTATGGAGTCATGCTAAATGAG 799
947 TGTGTTAGCTCTTGCTCTGGCAAAAAAAAAAAAAAAA 991
854 TAGATTTGGCTCTGGCACAGAACAAAAAATCTAAGGAA 754

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/csc>.

SOURCE INFORMATION:
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frangen, E., Tatano, M., Cataneo, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.rgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-35D5. 2000 bp overlap: the
MAMMALIA; *Eutheria*; **Primates**; *Catarrhini*; *Hominidae*; *Homo*.
 1 (bases 1 to 50630)

Clone sequenced to the right is RP11-642E20, 2000 bp overlap.
 Actual start of this clone is at base position 97704 of RP11-3SD5.

FEATURES source	Location/Qualifiers
repeat_region	/clone="RP11-400H10" /clone_id="RPCI-11" /rpt_family="Alu" /db_Xref="taxon:9606" /chromosome="4" /map="4" /rpt="RP11-400H10"
repeat_region	151..630 . Family="MalR" /rpt="RP11-3061" /rpt_family="L2" /rpt="RP11-3545" /rpt_family="L2" 1007..1305 . Family="Alu" /rpt="RP11-6142" /rpt_family="L1"
repeat_region	2780..3061 . Family="Alu" /rpt="RP11-7296" /rpt_family="L1" 6447..6446 . Family="Alu" /rpt="RP11-8521" /rpt_family="L1" 6143..6144 . Family="Alu" /rpt="RP11-8712" /rpt_family="L1" 7297..7669 . Family="Alu" /rpt="RP11-8767" /rpt_family="AT_rich" 7670..8151 . Family="Alu" /rpt="RP11-10937" /rpt_family="MIR" 8411..8521 . Family="Alu" /rpt="RP11-11244" /rpt_family="AT_rich" 10181..10668 . Family="L2" /rpt="RP11-12145" /rpt_family="L2" 10669..10936 . Family="Alu" /rpt="RP11-13466" /rpt_family="MIR" 14333..14336 . Family="Alu" /rpt="RP11-14395" /rpt_family="CAAA" 14677..14899 . Family="MIR" /rpt="RP11-15282" /rpt_family="MIR" 15282..15369 . Family="L2" /rpt="RP11-15487" /rpt_family="AT_rich" 15511..15572 . Family="Alu" /rpt="RP11-15607" /rpt_family="L2" 15882..16156 . Family="L1" /rpt="RP11-16164" /rpt_family="Alu" 16390..16406 . Family="MIR" /rpt="RP11-16465" /rpt_family="MER2_type" 16466..18051 . Family="Alu" /rpt="RP11-18052" /rpt_family="L1" 18370..18369 . Family="Alu" /rpt="RP11-18950" /rpt_family="L1" 18960..19059 . Family="Alu" /rpt="RP11-19231" /rpt_family="L1"
repeat_region	19322..20169 . Family="L1" /rpt="RP11-20170" /rpt_family="Alu" 20480..20630 . Family="L1" /rpt="RP11-20800" /rpt_family="MalR" 21687..21751 . Family="AT_rich" /rpt="RP11-21782" /rpt_family="L2" 23361..23671 . Family="Alu" /rpt="RP11-24505" /rpt_family="AT_rich" 26923..27223 . Family="Alu" /rpt="RP11-27248" /rpt_family="Alu" 27504..27930 . Family="Alu" /rpt="RP11-27951" /rpt_family="Tn" 28083..28152 . Family="Mariner" /rpt="RP11-28215" 29024..29147 . Family="AT_rich" /rpt="RP11-29148" /rpt_family="L2" 29148..29264 . Family="Alu" /rpt="RP11-29620" /rpt_family="Alu" 29620..29682 . Family="Mer1_type" /rpt="RP11-30174" /rpt_family="L2" 30570..30623 . Family="CRI1" /rpt="RP11-30632" /rpt_family="CR1" 30776..30921 . Family="AT_rich" /rpt="RP11-31991" /rpt_family="L1" 31420..33262 . Family="L1" /rpt="RP11-33396" /rpt_family="MIR"
Query Match	5..8t; Score 57..4; DB 9; Length 50630; Best Local Similarity 63..1t; Pred. No. 0..0041; Matches 123; Conservative 0; Mismatches 66; Indels 6; Gaps 2;
QY	793 AAGAGGGTTGGGGTTCAGGATAAGGAATGGGGATTGGAGCTGAGGACTGAGGAGCAG 852 Db 4996 AATGGTGTATAGGGCAATTAAATGGAGATGTCAGCCAGGGTACAAAGTAGCAG 4937
QY	953 CCTCTGAAATGAACTGAGGCAAAATGGAGATGTCAGCCAGGGTACAAAGTAGCAG 910 Db 4936 ATATGAGGATGAAAGTCTGGAGATTAATGACATGGGACTATGGTTAATAAA 4877
QY	911 GTGCACCTGCTATTGGAGTTCACTCAAATGAGTGTGTTAGCTGCTTGGCACA 970 Db 4876 TGTAA---TTCTATGAGGATCATGCTAAATGAGTGTGTTGGCACA 4821
QY	971 AAAAAAAAAAAAAA 985 Db 4820 AAAAAACACAAAAA 4806
RESULT 43	
AL357129	AL357129 Locus Human DNA sequence from clone RP11-112A21 on chromosome XG21-31-22.1 Contains ESTs, STSS and GSs, complete sequence.
DEFINITION	
XG21-31-22.1	
ACCESSION	AL357129
VERSION	AL357129..11 GI:10186549

KEYWORDS HTG. Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

REFERENCE Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1. (bases 1 to 145380) Chapman J.

TITLE Direct Submission

JOURNAL Submitted (19-Dec-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail: enquiries@sanger.ac.uk Clone requests: clone-requests@sanger.ac.uk

COMMENT On Sep 19, 2000 this sequence version replaced gi:9909073.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wo:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human Chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unseq' feature key. RP11-114A21 is from the library RP11-11 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: PBACE3_6

IMPORTANT: This sequence is not the entire insert of clone RP11-114A21. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-114A21 is at 1 in this sequence. The true left end of clone RP11-122K23 is at 145281 in this sequence. The true left end of clone RP11-554A24 is at 80790 in this sequence. The true right end of clone RP11-361B11 is at 54224 in this sequence.

FEATURES

source

1. 145380

/organism="Homo sapiens"
 /mol_type="Genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q21.31-22.1"
 /clone="RP11-114A21"
 /clone_lis="RP11-11.1"

1. .71

/note="LIP3 repeat: matches 10..80 of consensus"

20..685

/note="match: GSS: Em:A0343933"

31..79

/note="LIP repeat: matches 4..52 of consensus"

450..492

/note="MRE63 repeat: matches 2..44 of consensus"

847..993

/note="MRE20 repeat: matches 29..174 of consensus"

1591..1614

/note="12 copies 2 mer tc 95% conserved"

3441..3588

/note="2 copies 74 mer 92% conserved"

3742..3801

/note="L2 repeat: matches 2643..2701 of consensus"

4195..4440

/note="MUT1A1 repeat: matches 118..365 of consensus"

4442..4495

/note="HUERS_P3b repeat: matches 2227..2560 of consensus"

/note="AluSx repeat: matches 1..54 of consensus"

/note="HERV17 repeat: matches 1..1306..1630 of consensus"

repeat_region

4496..5265

/note="LIR7 repeat: matches 1..780 of consensus"

5271..6471

/note="LIPBa repeat: matches 1037..2339 of consensus"

6464..6757

/note="LIPBa repeat: matches -712..-411 of consensus"

6786..7118

/note="LIPBa repeat: matches -176..-1104 of consensus"

7770..8014

/note="MLTH repeat: matches 38..289 of consensus"

8097..8234

/note="MLTH repeat: matches 410..547 of consensus"

9501..10217

/note="LIP3 repeat: matches 4..924 of consensus"

8310..8458

/note="L2 repeat: matches 2563..2723 of consensus"

8484..8639

/note="L2 repeat: matches 1060..1224 of consensus"

10749..11059

/note="AluY repeat: matches 1..311 of consensus"

11101..11160

/note="LIP repeat: matches 948..2998 of consensus"

131159..15022

/note="LIP8 repeat: matches 4290..6162 of consensus"

15671..15793

/note="L2 repeat: matches 25771..2710 of consensus"

16891..17018

/note="MIR repeat: matches 14..144 of consensus"

18274..18889

/note="L2 repeat: matches 1656..2319 of consensus"

18868..18953

/note="MIR repeat: matches 169..254 of consensus"

19189..191961

/note="LIP2 repeat: matches 5536..6308 of consensus"

19962..21580

/note="LIP13 repeat: matches 4433..6073 of consensus"

21583..24399

/note="LIP13 repeat: matches 1532..4323 of consensus"

24420..25728

/note="LIP1 repeat: matches -301..1172 of consensus"

25721..25834

/note="LIM1 repeat: matches -1389..-1269 of consensus"

26112..26285

/note="MRE58A repeat: matches 45..224 of consensus"

26387..26742

/note="MIR1A1 repeat: matches 1..365 of consensus"

26857..27201

/note="L1ME repeat: matches 5514..5867 of consensus"

27851..27978

/note="MRE47 repeat: matches 2215..2323 of consensus"

27979..28352

/note="THE1C repeat: matches 5..371 of consensus"

28353..28503

/note="L1ME repeat: matches 2058..2215 of consensus"

28559..28645

/note="MIR repeat: matches 119..245 of consensus"

28672..29928

/note="MRE57-internal repeat: matches 1731..2937 of consensus"

29688..30062

/note="HUERS_P3b repeat: matches 2227..2560 of consensus"

30076..30409

/note="MRE57-internal repeat: matches 1306..1630 of consensus"

misc_feature

30550..30978

/note="match: GSS: Em:B3B708"

30554..30871

/note="match: GSS: Em:AQ589046"

31264..31806

/note="L2 repeat: matches 159..691 of consensus"

31816..32211

/note="AluSx repeat: matches 4..299 of consensus"

32245..32281

repeat_region	/note="PABL_B repeat: matches 204. .351 of consensus" 32282. .32586	LOCUS	AC021021	169202 bp	DNA	linear	PRI 07-NOV-2001
repeat_region	/note="PABL_B repeat: matches 7. .310 of consensus" 32587. .322734	DEFINITION	Homo sapiens BAC clone RP1-557N1	from 2,	complete	sequence.	
repeat_region	/note="PABL_B repeat: matches 102. .204 of consensus" 32735. .33047	ACCESSION	AC021021				
repeat_region	/note="AluSp repeat: matches 3. .313 of consensus" 33048. .33142	VERSION	1.9				
repeat_region	/note="PABL_B repeat: matches 1. .102 of consensus" 33207. .33283	KEYWORDS	GI:14140347				
misc_feature	/note="HUERS-P3b repeat: matches 7308. .7378 of consensus" complement(34673. .34083)	SOURCE	HTG.				
misc_feature	/note="match: GSS: Em:AQ034800" complement(34279. .34523)	ORGANISM	Homo sapiens (human)				
misc_feature	/note="match: STS: Em:558090 match: GSS: Em:AQ384354"	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.					
repeat_region	/note="match: GSS: Em:AZ664261"	REFERENCE					
repeat_region	/note="match: GSS: Em:AZ664261"	AUTHORS	VanBrunt,A., Hawkins,M., Haakenson,W. and Mullen,M.				
repeat_region	/note="HERVHC2 repeat: matches 4748. .5089 of consensus" 35211. .35546	TITLE	The sequence of Homo sapiens BAC clone RP1-557N1				
repeat_region	/note="HERVHC2 repeat: matches 4748. .5089 of consensus" 36577. .36598	JOURNAL	Unpublished				
repeat_region	/note="HUERS-P3 repeat: matches 4694. .5118 of consensus" 3698. .37624	REFERENCE	3 (bases 1 to 169202)				
repeat_region	/note="HRV repeat: matches 2672. .3008 of consensus" 38168. .38291.	AUTHORS	Waterston,R.H.				
repeat_region	/note="HRV9 repeat: matches 2732. .2840 of consensus" 38339. .38974	TITLE	Submitted (12-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
repeat_region	/note="YU2 repeat: matches 1. .36 of consensus" 40396. .40966	JOURNAL	4 (bases 1 to 169202)				
repeat_region	/note="PABL_B repeat: matches 1. .667 of consensus" 4076. .41346	REFERENCE	Waterston,R.H.				
repeat_region	/note="THE1C repeat: matches 15. .369 of consensus" 41348. .41495	AUTHORS	Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
repeat_region	/note="YER47 repeat: matches 2062. .2210 of consensus" 41525. .41641	TITLE	On May 17, 2001 this sequence version replaced g1:1339429.				
repeat_region	/note="YIR repeat: matches 119. .245 of consensus" 42737. .42875	JOURNAL	-----				
repeat_region	/note="YIR repeat: matches 110. .248 of consensus" 43217. .43703	COMMENT	-----				
repeat_region	/note="match: GSS: Em:AK522941"	REFERENCE	Center: Washington University Genome Sequencing Center				
repeat_region	/note="YI2 repeat: matches 2595. .2748 of consensus" 43929. .44074	AUTHORS	Center code: WIGSC				
repeat_region	/note="YI2 repeat: matches 110. .248 of consensus" 44087. .44550	TITLE	Web site: http://genome.wustl.edu/gsc				
repeat_region	/note="YIMCD repeat: matches 5354. .5823 of consensus" 44551. .44598	JOURNAL	Contact: sapiens@watson.wustl.edu				
repeat_region	/note="YIME1 repeat: matches 5685. .5825 of consensus" 44730. .44791	REFERENCE	-----				
repeat_region	/note="YIMD4 repeat: matches 6234. .6295 of consensus" 44794. .45004	AUTHORS	Summary Statistics				
repeat_region		TITLE	Center project name: H_NP055/N01				
misc_feature		JOURNAL	-----				
repeat_region		COMMENT	-----				
repeat_region		REFERENCE	-----				
repeat_region		AUTHORS	-----				
repeat_region		TITLE	-----				
repeat_region		JOURNAL	-----				
repeat_region		COMMENT	-----				
Query Match	Best Local Similarity 5.8%; Score 57.4; DB 9; Length 145380;	COMMENT	-----				
	Matches 109; Conservative 0; Mismatches 86; Indels 0; Gaps 0;	-----	-----				
Qy	795 GAGGAGTTGGGGGTTCAGGATAGGGAAATGGGAGGGTCAAGGAGCGAAAGCGAGGCC 854	-----	-----				
Dy	86358 GTGGGGATGGGAGGGACTAATGGAGATTAATGAGCTTCAAGACTAGTAGTATGAT 86417	-----	-----				
Qy	855 ATGAGAATGAACTGGCTTCAAGGAGGACTTGTTAGTGTGCTTAAATGAACTATGCTTAA 86477	MAPPING INFORMATION:	Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc				
Dy	86418 AAGTAGGATGAAATGAGCTTCAAGGAGCTTCAAGACTAGTAGTATGCTTAA 86477	-----	-----				
Qy	915 ACTGTTCTGATTGGAGTCTATGAAATGAGCTTCAAACTGCTTCAAGACTAGTAGTATGCTTAA 86537	SOURCE INFORMATION:	The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Prengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute				
Dy	86538 CACAAACAAGAAAA 86552	-----	-----				
RESULT	44	-----	-----				
	AC021021	-----	-----				

```
(http://bacpac.med.buffalo.edu)
VECTOR: PBACE_6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC021789; the clone sequenced to
the right is AC021688. Actual start of this clone is at base
position 1 of RP11-557N1; actual end is at base position 169202 of
RP11-557N1.

The sequence H NH0557N01 contains a dinucleotide (GA) repeat from
base position 29551 to 29868 where the sequence fidelity cannot be
guaranteed. Assembly of the database is consistent with digest
information.

FEATURES
source
repeat_region 13854..13946
/rept_family="Alu"
repeat_region 14020..14183
/rept_family="MIR"
repeat_region 14531..14841
/rept_family="Alu"
repeat_region 14860..15243
/rept_family="MIR2_type"
repeat_region 16037..16338
/rept_family="Alu"
repeat_region 17440..17564
/rept_family="CRI"
repeat_region 19909..19941
/rept_family="AT_rich"
repeat_region 20077..20139
/rept_family="MIR"
repeat_region 20366..20982
/rept_family="ERV1"
repeat_region 21156..21524
/rept_family="MAlR"
repeat_region 22036..22097
/rept_family="MIR"
repeat_region 22220..22331
/rept_family="MIR"
repeat_region 24135..24157
/rept_family="AT_rich"
repeat_region 25907..25968
/rept_family="MIR"
repeat_region 27874..28003
/rept_family="MIR"
repeat_region 28392..28426
/rept_family="MIR"
repeat_region 28637..28718
/rept_family=L2"
repeat_region 28774..28835
/rept_family="TTAAA)n"
repeat_region 29147..29261
/rept_family="GA-rich"
repeat_region 29287..29567
/rept_family="Alu"
repeat_region 29568..29642
/rept_family="polypurine"
repeat_region 29644..29831
/rept_family="GA-rich"
repeat_region 29832..29866
/rept_family=(GGA)n"
repeat_region 29991..30283
/rept_family="Alu"
repeat_region 30686..31803
/rept_family=L1"
repeat_region 31804..32105
/rept_family="Alu"
repeat_region 32106..32750
/rept_family="L1"
repeat_region 983 AAAA 989

Query Match 5.8%; Score 57.4; DB 9; Length 169202;
Best Local Similarity 64.2%; Pred. No. 0, 004;
Matches 120; Conservative 0; Mismatches 61; Indels 6; Gaps 2;

QY 805 GGGGTTCAGGATGGAAATGGGAGGTCAGAGGACGCCATGAGATG 864
Db 79581 GTGGAGAGGAACCTGGGGTAGGTAGGGATAAAATAGGTG 79640
QY 865 AACGGTCCAGAGCC-AAGCACGGGAGGACTGGGGCATCTAGGTGCG 922
Db 79641 AACAGGCTGGAGACCTAACCCACTAACGAGTAATAACTCGAGCT---G 79696
QY 923 TATTGGGACTCTGGAAATAAGTGTGTTTAAGCTGCTCTGGCA 982
Db 79697 TACCTGGGATTCTAGCTGAAATGCTGTTAGCTGCTCTGGCA 79756
QY 983 AAAA 989
```


MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McHerron, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Ooogaawa, K., Woon, P.Y., Zhao, B., Freijgen, E., Taterno, M., Cataneese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC-E6
NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-90C9, 2000 bp overlap; the clone sequenced to the right is RP11-22H19, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-469G18; actual end is at base position 1496 of RP11-22L19.

Data from AC079146 was used to finish this clone, AC013408. Polymorphisms have been identified between AC079146 and AC013408.

Location/Qualifiers

Source	Attributes
1..151841	/organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="ITaxon:9606" /chromosome="2" /map="2"
	/clone="RP11-469G18" /cclone_id="RP11-11" /clone_start="1209" /clone_end="1210..2222" /rept_family="ERVK"
repeat_region	/rpt_family="MIR"
repeat_region	/rpt_family="L1"
repeat_region	3559..3553
repeat_region	/rpt_family="(CA)n"
repeat_region	5300..5476
repeat_region	/rpt_family="MIR"
repeat_region	5483..5601
repeat_region	/rpt_family="AChObo"
repeat_region	6149..6256
repeat_region	/rpt_family="L12"
repeat_region	6526..6700
repeat_region	/rpt_family="MAlR"
repeat_region	9017..9184
repeat_region	/rpt_family="(TA)n"
repeat_region	9834..9864
repeat_region	/rpt_family="(GA)n"
repeat_region	12527..12597
repeat_region	/rpt_family="TA)n"
repeat_region	13139..13760
repeat_region	/rpt_family="AT-rich"
repeat_region	13964..13999
repeat_region	/rpt_family="Alu"
repeat_region	14037..14242
repeat_region	/rpt_family="MAlR"
repeat_region	14244..14240
repeat_region	/rpt_family="(TC)n"
repeat_region	14833..14787
repeat_region	/rpt_family="Alu"
repeat_region	15016..15042
repeat_region	/rpt_family="(TAGA)n"
repeat_region	15169..15555
repeat_region	/rpt_family="Alu"
repeat_region	17126..17147
repeat_region	/rpt_family="(TTTTA)n"
repeat_region	17131..17430

Query Match Similarity 5.7% Score 56.6 Best Local Match 56.0% Pred. No.: March 10, 2005

TITLE Direct Submission
 JOURNAL 2 (bases 1 to 164405)
 AUTHORS Worley K.C.
 JOURNAL Direct Submission
 Submitted (24-NOV-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 164405)
 AUTHORS Worley K.C.
 JOURNAL Direct Submission (11-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 164405)
 AUTHORS Worley K.C.
 JOURNAL Direct Submission (25-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2001 this sequence version replaced gi:12656764.
 INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
 sc-help@bcm.tmc.edu

CLOSE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
 STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
 Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
 Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect <1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES source
 Location/Qualifiers
 1.
 .164405
 /organism="Homo sapiens"
 /mol_type="Genomic DNA"
 /db_xref="taxon:2606"
 /chromosome="12"
 /clone="RP11-98E6"
 repeat_region
 misc_feature
 3.
 .5166
 /rpt_family="MLT1F1"
 /note="Overlaps bases 186760. 19193 of clone AC087886"
 complement(439. 637)
 /rpt_family="LIMA8"

repeat_region
 2142. .2281.
 /rpt_family="MIR"
 complement(2827. 2961)
 /rpt_family="MIR"
 3943. .4314.
 repeat_region
 5502. .5527
 /rpt_family="AT_rich"
 6069. .6115
 /rpt_family="LIME3A"
 6830. .6111
 /rpt_family="L2"
 repeat_region
 7015. .7068
 /rpt_family="AT_rich"
 7069. .7053
 /rpt_family="(TA)n"
 7163. .7191
 repeat_region
 complement(7132. 7324)
 /rpt_family="L2"
 repeat_region
 7825. .8933
 complement(7387. 7463)
 /rpt_family="L1PA13"
 7825. .8933
 complement(9397. 9171)
 /rpt_family="L2"
 repeat_region
 949. .9629
 /rpt_family="L1"
 repeat_region
 10071. .10075
 /rpt_family="LIMA9"
 10394. .10581
 /rpt_family="L1M4"
 repeat_region
 10582. .11009
 /rpt_family="Trigger2a"
 repeat_region
 11010. .11303
 /rpt_family="L1M4"
 11500. .11670
 /rpt_family="L1M3"
 repeat_region
 11701. .12100
 /rpt_family="L1M3"
 12185. .13027
 /rpt_family="L1M3"
 repeat_region
 complement(13224. 13505)
 /rpt_family="L1PA10"
 13929. .133957
 /rpt_family="AT_rich"
 repeat_region
 14216. .15549
 misc_feature
 /rpt_family="L1PA3"
 /note="C or A"
 /function="Polymorphic site"
 misc_feature
 14288
 /note="C or T"
 /function="Polymorphic site"
 misc_feature
 STS
 14300. .14431
 /standard_name="G19048"
 /db_xref="dbSTS:32886"
 misc_feature
 14356
 /note="C or T"
 /function="Polymorphic site"
 misc_feature
 14362
 /note="C or G"
 /function="Polymorphic site"
 misc_feature
 14397. .14398
 /note="CA or TG"

Db 111101 GCTAAATGAGTCGATTTAGCTGCTCTGCCACAAACCAAAAAAAA 111046 misc_feature 1. .26915
 RESULT 50 /note="assembly_name:Contig8
 AC011004/c LOCUS AC011004 183778 bp DNA linear HTG 11-MAR-2001 clone_end:SP6" 27016-.98873
 DEFINITION Homo sapiens chromosome 8 clone RP11-110D15, WORKING DRAFT /note="assembly_name:Contig9
 ACCESSION AC011004 3 unordered pieces. clone_end:T7" 98974-.183778
 KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP /note="assembly_name:Contig10"
 SOURCE Homo sapiens (human) BASE COUNT 50361 a 41935 c 41401 g 49880 t 201 others
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1. (bases 1 to 183778) misc_feature 5. 6%; Score 56; DB 2; Length 183778;
 AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R., Komp,C., Kotlikoff,S., Lam,B., Marath,R., Miranda,M., Morhouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W. Matches 116; Conservative 0; Mismatches 60; Indels 4; Gaps 2;
 JOURNAL Unpublished QY 813 AGGATAGGGAAATGGGGAGGTCAAGGAGCCAAAGGAGCAGCCATGAGATGAACCGTCC 872
 REFERENCE 2. (bases 1 to 183778) Db 37638 AGGAACGGGAGATAGGGTGAAGAACGAGAGGAGATATCAGATGATGCGCC 37577
 AUTHORS Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P., Glukhov,S., Hansen,N., Hyman,R., Marath,R., Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D. and Davis,R.W. QY 873 AGAGACCCAAACGACGCCAGGGACTGAGGCCATAGCGGTGCACTTGCGTATTTGGAGT 932
 TITLE Submitted (29-SEP-1999) DNA Sequencing and Technology Center, Db 37578 AGAGACCTAATGT-SCATGAGGACTCTACTCAAGACATGTA 37523
 JOURNAL Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
 COMMENT On Mar 11, 2001 this sequence version replaced gi:13270571. Search completed: February 9, 2004, 14:16:33
 Center: Stanford DNA Sequencing and Technology Development
 Center Center code: SDSTDC Job time : 3928 secs
 Web site: http://sequence-www.stanford.edu/group/human/
 Contact: hum-infouse@stanford.edu
 ----- Project Information
 Center clone name: RP11-110D15
 Center project name: 666
 ----- Summary Statistics
 Sequencing Vector: M1mp18; X02513
 Chemistry: Dye-primer; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.90319
 Consensus quality: 182108 bases at least Q40
 Consensus quality: 183015 bases at least Q30
 Consensus quality: 183336 bases at least Q20
 Insert size: 185712; agarose-fp
 Insert size: 183578; sum-of-contigs
 Quality coverage: 9.3x in Q20 bases; agarose-fp
 Quality coverage: 9.4x in Q20 bases; sum-of-contigs.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence,
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 26915: Contig of 26915 bp in length
 * 26916 27015: Gap of unknown length
 * 27016 98873: Contig of 71858 bp in length
 * 98874 98973: Gap of unknown length
 * 98974 183778: Contig of 84805 bp in length.
 FEATURES Location/Qualifiers
 source 1..183778
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="18"
 /clone="RP11-110D15"
 /clone_lib="RPCI human BAC library 11"

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 13:03:52 : Search time 2524 Seconds
(without alignment)

Title: US-09-990-726-222
Perfect score: 992
Sequence: 1 ggcacgaggccggaaacttgg.....aaaaaaaaaaaaaaa 992

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing First 150 summaries

Database : EST:*

Result No.	Score	Match Length	DB ID	Description
1	626.8	63.2	694 13 BX097129	BX097129 BX097129
c 2	409.8	41.3	414 14 H74303	H74303 yu56cl1.s1
c 3	408	41.1	437 14 NS4458	NS4458 yv40b02.s1
c 4	373.2	37.6	480 14 H73373	H73373 yu48f10.r1
5	354	35.7		
	6	349	35.2	H59326
	7	340.6	34.3	H4302
	8	320.4	32.3	N76568
c 9	286	28.8		AK020081
c 10	284.2	28.6		A1242058 qhd1d07.X
c 11	277.2	27.9		H73374 Yu4f10.s1
c 12	274.2	27.6		F02548 yeBa07.s1
c 13	268.8	27.1		A143986 tc4aa07.s1
c 14	265.2	26.7		BY718164 BY718164
c 15	261.2	26.3		B+529784 RZD MuS
c 16	256.8	25.9		H40623 yp5e11.s1
c 17	235.2	23.7		H58715 Yr2c07.s1
c 18	227.4	22.9		TB0187 yd8sg03.s1
c 19	220.2	22.2		AP67653 zj6re03.s
c 20	218.4	22.0		BY229784
c 21	218.4	22.0		R026666 yeBa07.r1
c 22	202	20.4		WA1959 mc6bb10.r1
c 23	199.2	19.4		AP050439 m720f02.r
c 24	182.2	18.4		TB7491 yd8sg03.r1
c 25	175.6	17.7		WB5713 me09706.r1
c 26	174.6	17.6		NB6939 L1498F Huma
c 27	154	15.5		BY212485 BY212485
c 28	153.8	15.5		BI347002
c 29	153.8	15.5		BI347003
c 30	153.8	15.5		CB051335 NISC_g123
c 31	145.2	14.6		CB051334 NISC_g123
c 32	143	14.4		NB8474 K375F Huma
c 33	132.6	13.4		BB842514 BB842514
c 34	93.2	9.4		BB843270 BB843270
c 35	92.6	9.3		BF443482 260980 MA
c 36	79.8	8.0		BF442092 258654 MA
c 37	57.8	5.8		AO31034 mi7205.r1
c 38	52.8	5.3		AI187208 GT2B11.x
c 39	52.6	5.3		AW794912 RC6-UM001
c 40	51.2	5.2		Bx487006 Homo sapi
c 41	50.8	5.1		BI1063151 IL3-UT011
c 42	4.9	4.9		BB772737 mcV5102.
c 43	4.8	4.8		CB615918 AMGNNUC.U
c 44	47.4	4.8		CD39738 Gm_ck1771
c 45	47.2	4.8		AL066051 Drosophil
c 46	46.4	4.7		BX339244 BX339244
c 47	46.2	4.7		AL053013 Drosophil
c 48	45.6	4.6		BZ772737 mcV5102.
c 49	45.6	4.6		CB615664 BX356664
c 50	45.6	4.6		CD39738 Gm_ck1771
c 51	45.4	4.6		AL066051 Drosophil
c 52	45.4	4.6		BX339244 BX339244
c 53	45	4.5		AL053013 Drosophil
c 54	44.8	4.5		BZ772737 mcV5102.
c 55	44.6	4.5		CB615918 AMGNNUC.U
c 56	44.2	4.5		CD39738 Gm_ck1771
c 57	43.4	4.4		AL066051 Drosophil
c 58	43.2	4.4		BX339244 BX339244
c 59	43	4.3		AL053013 Drosophil
c 60	43	4.3		BZ772737 mcV5102.
c 61	43	4.3		CB615918 AMGNNUC.U
c 62	43	4.3		CD39738 Gm_ck1771
c 63	42.6	4.3		AL066051 Drosophil
c 64	42.6	4.3		BX339244 BX339244
c 65	42.6	4.3		AL053013 Drosophil
c 66	42.6	4.3		BZ772737 mcV5102.
c 67	42.6	4.3		CB615918 AMGNNUC.U
c 68	42.6	4.3		CD39738 Gm_ck1771
c 69	42.6	4.3		AL066051 Drosophil
c 70	42.4	4.3		BX339244 BX339244
c 71	42.2	4.3		AL053013 Drosophil
c 72	42.2	4.3		BZ772737 mcV5102.
c 73	42.2	4.3		CB615918 AMGNNUC.U
c 74	42	4.3		CD39738 Gm_ck1771
c 75	42	4.2		AL066051 Drosophil
c 76	41.8	4.2		BX339244 BX339244
c 77	41.8	4.2		AL053013 Drosophil

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Query	Match Length	DB ID	Description
AI43986 tc4aa07.s1	13	BX097129	BX097129 BX097129
H74303 yu56cl1.s1	14	H74303	H74303 yu56cl1.s1
NS4458 yv40b02.s1	14	NS4458	NS4458 yv40b02.s1
H73373 yu48f10.r1	14	H73373	H73373 yu48f10.r1

Result No.	Score	Match Length	DB ID	Description
1	626.8	63.2	694 13 BX097129	BX097129 BX097129
c 2	409.8	41.3	414 14 H74303	H74303 yu56cl1.s1
c 3	408	41.1	437 14 NS4458	NS4458 yv40b02.s1
c 4	373.2	37.6	480 14 H73373	H73373 yu48f10.r1

Result No.	Score	Match Length	DB ID	Description
1	626.8	63.2	694 13 BX097129	BX097129 BX097129
c 2	409.8	41.3	414 14 H74303	H74303 yu56cl1.s1
c 3	408	41.1	437 14 NS4458	NS4458 yv40b02.s1
c 4	373.2	37.6	480 14 H73373	H73373 yu48f10.r1

Insert Size: 1114
 High quality sequence stops: 313
 Source: IMAGE Consortium, LInL
 This clone is available royalty-free through LInL ; contact the
 IMAGE Consortium (info@image-linl.gov) for further information.
 Insert Length: 1114 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence stop: 313.
 Location/Qualifiers
 1. .414
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:3781228"
 /db_xref="taxon:9606"
 /clone="IMAG3:230132"
 /sex="male"
 /dev_stage="20 week"-post conception fetus"
 /clone_lib="Soares fetal liver spleen INFSL"
 /note="Organ: Liver and Spleen; Vector: pRT73D (Pharmacia)
 with a modified polylinker; Site 1: site I - oligo(dT) primer
 [5'- AACATCGAACTATTAAGATCTTTTTTTTTTTTTTTTT 3'] ,
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pRT3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 116 g 95 t 1 others

Db	FEATURES	source
403	GCGCTAAGATGCACTGACTGGAGCTCATCTGCCAGCCAAACGGCAAGCTGCCTGGCCA 462	
185	GCGCTAAGATGCACTGAGGCTGTGTC 214	
463	CTCTGCAGGAGAGGGCAGGGGAGATCATCTGCCAGCCAAACGGCAAGCTGCCTGGCCA 522	
215	-----GAGAGGGGAGGCGCCAGGGGAGATCATCTGCCAGCCAAACGGCAAGCTGCCTGGCCA 266	
523	GCCCCCTATCACAAAGCGCTGATCGGGAGGATCATCTGCCAGCCAAACGGCAAGAGAC 582	
267	-----GAGAGGGGAGGCGCCAGGGGAGATCATCTGCCAGCCAAACGGCAAGCTGCCTGGCCA 326	
583	CATGCCACAGGAGGCGCTGCGCAACTCTCGCTGCCAGAGAACATCGAGCTGGTCT 642	
327	CATGCCACAGGAGGCGCTGCGCAACTCTCGCTGCCAGAGAACATCGAGCTGGTCT 386	
643	GGTGCAGGCTGAAACAAGCAATGCTCCAGACACGGCCCTCACAGTGGTSCCCCCAG 702	
387	GGTGCAGGCTGCAACAAACGCCAACGCGCCAAATGCTCCAGACACGGCCCTCACAGTGGTSCCCCCAG 446	
703	GTTGGTACCCAGAACATGGGAGACTGGCAGGCTGGCAGGCTGGCAGGCTGGCTTGC 762	
447	GTTGGTACCCAGAACATGGGAGACTGGCAGGCTGGCAGGCTGGCAGGCTGGCTTGC 506	
763	GCGCTAAGAGGAGGACCCGGCTCAAGTGAGGAGGTTGGGGGTTCAAGGATAGGA 822	
507	GCGCTAAGAGGAGGACCCGGCTCAAGTGAGGAGGTTGGGGGTTCAAGGATAGGA 566	
823	AT-GGGGAGTCAAGGAGCCATCTAGATGAAACGGTCCAGAGGCCA 881	
567	ATGGGGAGGTTCAAGGAGCCATCTAGATGAAACGGTCCAGAGGCCA 626	
2Y	AGCAAGGGCAAGGGACTGAGCCATAGCGCTGCTATTGGAGTTCTCATGCAA 941	
Db	AGCAAGGGCAAGGGACTGAGCCATAGCGCTGCTATTGGAGTTCTCATGCAA 686	
942	ATGGATGT 949	
687	ATGGATGT 694	
2Y	414 bp mRNA linear EST 31-OCT-1995	
H74303	YU56c11.s1 Soares fetal liver spleen INFSL Homo sapiens cDNA clone IMAGE:230132 3', mRNA sequence.	
DEFINITION	H74303.1 GI:1047714	
ACCESSION	EST.	
KEYWORDS		
ORGANISM	Homo sapiens (human)	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
Hillier,L., Lemion,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,		
Chissoe,S., Dietrich,N., DuBuque,T., Faville,A., Gish,W., Hawkins		
M., Hullman,M., Kucaba,T., Lacy,M., Le,N., Le,N., Moore		
,B., Morris,M., Parsons,J., Prange,C., Rafkin,L., Rohlfing,T.,		
Schellenberg,K., Soares,M.B., Tan,F., Thierry,Meg,J., Trevaskis,E.,		
Underwood,K., Wohldmann,P., Waterston,R., Wilson,R., and Marra,M.		
Generation and analysis of 280,000 human expressed sequence tags		
Genome Res. 6 (9), 807-828 (1996)		
RESULT	3	
N54458/C		
LOCUS	YV0B02..s1 Soares fetal liver spleen INFSL Homo sapiens cDNA clone	
DEFINITION	IMAGE:245163 3', mRNA sequence.	
ACCESSION	N54458	
VERSION	EST.	
KEYWORDS	Homo sapiens (human)	
SOURCE		
TITLE		
MEDLINE		
PUBMED	8889549	
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286-1800 Fax: 314 286-1800 Email: est@wustl.edu	
RESULT	3	
N54458		
LOCUS	437 bp mRNA	
DEFINITION	linear EST 28-JAN-1997	
IMAGE	YV0B02..s1 Soares fetal liver spleen INFSL Homo sapiens cDNA clone	
ACCESSION	N54458	
VERSION	EST.	
KEYWORDS		
SOURCE		
TITLE		
MEDLINE		

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
(bases 1 to 437)	
REFERENCE	137 AGGAGGAAACAGGCCATTGATGAGTGAATGTAGAGAGCCAGGGAGGA 78
AUTHORS	QY 896 CTGCAGGCCATTGCTGCACTGTGTTATGGACTTCATGCAAATGACTGTGTTTA 955
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiappelli,B.,	Db 77 CTGCAGGCCATTGCTGCACTGTGACTGTGTTATGGACTTCATGCAAATGACTGTGTTTA 18
Chissoe,S., Dietrich,N., DuBque,T., Favello,A., Gish,W., Hawkins	
'B., Hullman,M., Kucaba,T., Lacy,M., Le,N., Mardis,E., Moore	
'B., Morris,M., Parsons,J., Riekin,L., Rohlfing,T., Trevaskiss,E.,	
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.	
TITLE	QY 956 GCTGCTCTTGCCACAA 972
JOURNAL	Db 17 GCTGCTCTTGCCACAA 1
MEDLINE	
PUBMED	
COMMENT	
Contract: Wilson RK	
Washington University School of Medicine	
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108	
Tel: 314 286 1800	
Fax: 314 286 1810	
Email: est@wtソン.wustl.edu	
This clone is available royalty-free through LInL ; contact the	
IMAGE Consortium (info@image.lnl.gov) for further information.	
Insert Length: 832 Std Error: 0.00	
Seq primer: ml3_40 Forward	
High quality sequence stop: 375.	
Location/Qualifiers	
1..437	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="GDB:3794409"	
/db_xref="taxon:9006"	
/clone="IMAGB:245163"	
/sex="male"	
/dev_stage="20 week-post conception fetus"	
/clone_lib="Soares fetal liver spleen INFSL"	
/note="Organ: Liver and Spleen; Vector: pRT3D (Pharmacia)	
with a modified Polylinker; Site_1: Pac I; Site_2: Eco RI;	
1st strand cDNA was primed with a Pac I - Oligo(dT) primer	
[5'-AACNGAAATTAAATTAAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3'] ,	
double-stranded cDNA was ligated to Eco RI adaptors	
(Pharmacia), digested with Pac I and cloned into the Pac I	
and Eco RI sites of the modified pRT3D vector. Library	
went through one round of normalization. Library	
constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	78 a 132 c 119 g 106 t 2 others
ORIGIN	
Query Match	41.1%
Best Local Similarity	97.3%
Matches	425; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
FEATURES	
source	1..480
Location/Qualifiers	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="GDB:3780483"	
/db_xref="taxon:9606"	
/clone="IMAG3:229387"	
/sex="male"	
/dev_stage="20 week-post conception fetus"	
/clone_lib="Soares fetal liver spleen INFSL"	
/note="Organ: Liver and Spleen; Vector: pRT3D (Pharmacia)	
1st strand cDNA was primed with a Pac I - oligo(dT) primer	
[5'-AACNGAAATTAAATTAAGATTTTTTTTTTTTTTTTTTTTTTTTTTTT 3'] ,	
double-stranded cDNA was ligated to Eco RI adaptors	
(Pharmacia), digested with Pac I and cloned into the Pac I	
and Eco RI sites of the modified pRT3D vector. Library	
went through one round of normalization. Library	
constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	100 a 149 c 130 g 100 t 1 others
ORIGIN	
Query Match	37.6%; Score 373.2; DB 14; Length 480;

MEDLINE PURNED	97044478 889549	Contract: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu	High quality sequence stops: 308 Source: IMAGE Consortium, LiNL This clone is available royalty-free through LiNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1114 Std Error: 0.00 Seq primer: M13RPI High quality sequence stop: 308. Location/Qualifiers	421 AGTTTGTGTTCAAAGGCCAGTGTNTGAGGTTGGGCAATTNNAAATTGAGGAGG 480
FEATURES	source			
	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:3781228" /db_xref="taxon:9606" /clone_id="IMAGE:230132" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DHIOB (ampicillin resistant)" /clone_lib="Soares fetal liver spleen INFLS" /note="Organ: Liver and Spleen; Vector: pRT3D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGGAAAGATTAATAGCTTTCATTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pRT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."			
BASE COUNT	126 a 149 c 157 g 130 t 20 others			
ORIGIN				
Query Match	35.2%	Score 349, DB 14; Length 582;		
Best Local Similarity	87.0% ; Matches 470 ; Conservative 0 ; Mismatches 53 ; Indels 17 ; Gaps 8 ;			
Features	source			
	10 CAGGAACCTGAGGTCTACTGCCATCAAAGCTGCTGGCTGCAGC- 1 CAGGAACCTGAGGTCTACTGCCAGCAAGCCCTAACCCAGG 60			
	69 CTCCCTGGGTGTTGCTGCCCTGCTGCAGC- 61 CTCCCTGGGTGTTGCTGCCCTGCTGCAGC- 128 GAGAAATTACCCCTGTTCTCATGCCATCAAAGCTGCTGGCTGCAGCAGC 127			
	121 GAGAAATTACCCCTGTTCTCATGCCATCAAAGCTGCTGGCTGCAGCAGC 180			
	188 CGCTGGCTCATAACCTGCTGTGACCCACGCCATCACTTATCCCT 247			
	181 CGCTGGCTCATAACCTGCTGTGACCCACGCCATCACTTATCCCT 240			
	248 CTGGAAACCAAGAACATCAAGTGCGCAAGAACGTTGCAAGACGCCCTC 307			
	241 CTGGAAACCAAGAACATCAAGTGCGCAAGAACGTTGCAAGACGCCCTC 300			
	308 CTGGAAACCTCAAGCTCACACTCAAGTCCGTCCGACCT-GCTACCTACT- 301 CTGGAACTCAACCTCACACTCAAGTCCGTCCGACCTACTATGGCG 360			
	365 GGCTCTCCATCCCTGCGCCATGTGG-- ACAGTGCAGGCTACAGATGCACTGG 421			
	361 GGCTCTCCATCCCTGCGCCATGTGGACAGTGTCAAGATGCACTGG 420			
	422 GCGCTT-GGTCCAAGGCCAGTGTGTGAGCTGAGGAGCAG 475			
BASE COUNT	105 a 156 c 126 g 80 t 1 others			
ORIGIN				
Query Match	34.3%	Score 340.6 ; DB 14 ; Length 468 ;		
Best Local Similarity	90.3% ; Matches 457 ; Conservative 0 ; Mismatches 5 ; Indels 44 ; Gaps 7 ;			
	218 CCAGGCCAACCCGCCATCACCTATCCCTGTGGANCAAGACATCAAGTGGCGCA 277			
	2 CCAGGCCAACCCGCCATCACCTATCCCTGTGGANCAAGACATCAAGTGGCGCA 61			
ORIGIN				
Query Match	34.3%	Score 340.6 ; DB 14 ; Length 468 ;		
Best Local Similarity	90.3% ; Matches 457 ; Conservative 0 ; Mismatches 5 ; Indels 44 ; Gaps 7 ;			
	218 CCAGGCCAACCCGCCATCACCTATCCCTGTGGANCAAGACATCAAGTGGCGCA 277			
	2 CCAGGCCAACCCGCCATCACCTATCCCTGTGGANCAAGACATCAAGTGGCGCA 61			
ORIGIN				
Query Match	34.3%	Score 340.6 ; DB 14 ; Length 468 ;		
Best Local Similarity	90.3% ; Matches 457 ; Conservative 0 ; Mismatches 5 ; Indels 44 ; Gaps 7 ;			
	218 CCAGGCCAACCCGCCATCACCTATCCCTGTGGANCAAGACATCAAGTGGCGCA 277			
	2 CCAGGCCAACCCGCCATCACCTATCCCTGTGGANCAAGACATCAAGTGGCGCA 61			
ORIGIN				
Query Match	34.3%	Score 340.6 ; DB 14 ; Length 468 ;		
Best Local Similarity	90.3% ; Matches 457 ; Conservative 0 ; Mismatches 5 ; Indels 44 ; Gaps 7 ;			
	218 CCAGGCCAACCCGCCATCACCTATCCCTGTGGANCAAGACATCAAGTGGCGCA 277			
	2 CCAGGCCAACCCGCCATCACCTATCCCTGTGGANCAAGACATCAAGTGGCGCA 61			
ORIGIN				

/db_xref="MGI:1911589"						
/db_xref="taxon:10030"						
/clone="6030468B19"						
/sex="male"						
/tissue_type="testis"						
/clone["Riken full-length enriched mouse cDNA library"]						
/dev_stage="13 days embryo"						
33. - 791						
CDS						
/note="unnamed protein product; hypothetical putative"						
Immunoglobulin structure containing protein (SCOP 48726, evidence: SCOP)						
/codon_start=1						
/protein_id="BAB31985.1"						
/db_xref="IGI:12860551"						
/db_xref="MGI:1924977"						
/translation="MALLQILFAMLAACGFSSEBQETTAKVYKLEVYPOSSRRVLIT						
CDAPASOPTYSVSLAASPGTVALAKVYDVSPPSNTINITISPDLLTSCAT						
GTYGPFSSRLQMYQELWAKPSQLADFLVRHGSPTVLSCLASSGSPPTIVGN						
GGRLAQALPFLPHGRPANFLPLSPOTTGWFQCEAENDVGVDSSARIPLRAEARAKLVT						
TIALEBLPLPTTLAGLSVSYIASVARMUSSLTGIL"						
977. - 982						
/note="putative"						
995						
polyA_signal						
polyA_site						
BASE COUNT	217	a	312	c	257	g
ORIGIN						
Query Match	32.3%		Score 320 4;	DB 11;	Length 996;	
Best Local Similarity	68.9%		Pred. No. 6e-36;			
Matches 460;	Conservative 0;	Mismatches 196;	Indels 12;	Gaps 1;		
Db	14	GCCCCAACCTTCAGAGCTCACAGACGATGGCCTCCCTCAGGTGCGTC	103			
Qy	44	GCCCTACACCACCGAGGATGGGGCTCCTGGCTGTCTGCTGGCTGC	103			
Db	14	-CCATTGCTTCA	121			
Qy	104	CAGCGCTTCTCCAAGCAAGGCGTGGGTGCTCATACCTGCTGTCACCCAGCC	223			
Db	74	-CTGTCGCTTCCTCAGGGAGAGAACGGCTTCA	121			
Qy	164	AGTCCTGGAAGTTTCCCAAAGGGCGTGGGTGCTCATACCTGCTGTCACCCAGCC	223			
Db	122	-AGTACTGGAACTTATCCCAGGGAGAGAACGGCTTATAACCTGGAATCCCCTGAGGC	181			
Qy	224	ACACCGCCATCACCTATCCCTCTGTGAAACCAAGAACATCAAGGTGGCCAAGAAGT	283			
Db	182	GTCCCCAGCCATCACATCTCTCTGGTGGTAGGCCAGGTATCTGGTCCAAAAGT	241			
Qy	284	GGTGAAGACCAAGGCCGCGCTCTTCAACCTCACTGTAACCTCAAGTCAGTCAGTC	343			
Db	242	TGTGCTGATCTCCGTCGCGCTCCCTCACATCAATACATCAAGGCCAGCA	301			
Qy	344	CCTGTCACCTACTCTGCGGAGCTCTGCCCTCAGTCGGTGGACAGTCGCCAG	403			
Db	302	CTTGTCACTACTCTGCGGAGCTCTGCCCTCAGTCGGTGGACCTATGGACCATGCA	361			
Qy	404	GCTAAGATGACTCTGGAGCTGCTGTCAGCCACTGCTGAGCTGCGGCCAACTTCAC	463			
Db	362	GCTCCAGATGTACCGAAACTGTCGCTCACTGCTGAGCTGCACTCTGACTCTC	421			
Qy	464	TCTGCAAGGAACTGCTGCAACTGCTGCTGAGCTGCTGAGCTGCTGCGCTCTGGCG	523			
Db	422	CCTACCCCATGGGACTGCGCCCACTGTCGCTGAGCTGCTGAGCTGCTGAGCG	481			
Qy	524	CCACCTATCACAAAGCTGTCGAGGATGGGAGTGGAGCTGCCATCTGAGAGCC	583			
Db	482	CCCCCCCACATCACCTACCTACCTGCTGAGGAGTGGAGCTGCCATCTGAGAGCC	511			
Qy	584	ATGCCACGGCAGCCACTCTCCCTCTGCGGCCAGACATGGACTCTGGCTC	643			
Db	542	ACTCATGAAAACGCCAACCTCCCTCCGACCACTCTGCTGTTCA	601			
Qy	879	GTCGGAGGCTGCAAGGACTGCAAGGAGCTGCAAGGAGCTGCAAGGAGCTGCA	703			

FEATURES	source	COMMENT
		Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp; URL: http://genome.gsc.riken.go.jp/ Adachi,J., Aizawa,K., Hayashizaki,N., Akarawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashizaki,K., Hirozane,T., Hori,F., Imatori,K., Ishii,Y., Itch,M., Kasuga,I., Kawai,J., Koijima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohnishi,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Wakai,K., Watashiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
		Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing Pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
Location/Qualifiers		
1. 553		/organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="6030468B19" /sex="male" /tissue_type="testis" /dev_stage="13 days embryo" /lab_host="DH110B" /clone_lib="RIKEN full-length enriched, 13 days embryo male testis"
		/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAAGGATCCAGAGCTCTTTTCTTTTN 3']". cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTCGAGTAAATTAACTCCCTCCC 3']. cDNA was cloned into the XbaI and BamHI sites. Vector: a Lambda FIC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT	118 a	Query Match Score 268.8; DB 14; Length 553; Best Local Similarity 70.0%; Pred. No. 1..3e-28; Matches 382; Conservative 0; Mismatches 152; Indels 12; Gaps 1;
ORIGIN	137 g	44 GCCCTAACCCACCGAGGATGGGCTCCCTGGGCTGCTGCTGCCTC 103
		RZPD; IMAP998K041131; RZPDLIB; I.M.A.G.E.; cDNA Clone Collection; CC Mouse UniGeneSet - RZPDLIB No. 981) http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=981 Contact: Ina Rolfs

CC RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
 CC Heubnerweg 6, D-14059 Berlin, Germany
 CC Tel: +49 30 32639 101
 CC Fax: +49 30 32639 111
 CC www.rzpd.de
 CC This clone is available royalty-free from RZPD:
 CC contact RZPD (clone@rzpd.de) for further information.
 CC Seq primer: "T7, Primer sequence: TAATAGCTACTATAGG
 XX
 Key Location/Qualifiers
 FH source 1..528
 FT /db_xref="txaxon:10090"
 FT /note="1st strand cDNA was primed with a Not I - oligo (dT)
 FT primer [5' TGTATCCATTCTGAGTGGAGGCCGGAAATTTTTTT
 FT TTTTTTTTTTTTTTTT 3'], on equal amounts of mRNA from 2
 FT 13.5dpc and 14.5dpc embryos (total RNA provided by Minoru
 FT Ko, Wayne State Univ., From 2], double stranded cDNA was
 FT ligated to EcoRI adaptors 5'-AATTGGCAAGAGG-3' and
 FT 5'-CCMCCTGCCG-3' (Pharmacia), digested with NotI and cloned
 FT into the NotI and EcoRI sites of the pRPT3D-Paci vector.
 FT Library went through one round of normalization, and was
 FT constructed by Bento Soares and M. Fatima Bonaldo."
 FT /organism="Mus musculus"
 FT /clone_id="INAGp98K041131"
 FT /clone_lib="Soares 13.5-14.5dpc embryo NbMME13.5-14.5"
 FT /dev_stage="13.5-14.5dpc total fetus"
 FT /lab_host="DH10B"
 XX Sequence 528 BP; 112 A; 177 C; 132 G; 107 T; 0 other;
 SQ

Query Match 26.7%; Score 265.2%; DB 4; Length 528;
 Best Local Similarity 71.6%; Pred. No. 4.e-28;
 Matches 348; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Source	FEATURES
Qy	148 TCTCCATGCTTACAAAGTCTGGAAAGTTTCCCANAGGCCTCTTCAACCT 207
Db	42 TCACCATGCTTACAAAGTACTGGAAAGTTTCACCTTCCTCTGGTATAACCT 101
Qy	208 GCTGTGCACCCAGCCACCACCCCCATACCTTCCCTCTGGAACAAAGAACATCA 267
Db	102 GCGATGCCCTGAAGGGTCCAGCCATACATCTCTCTGGTAGCCAGGTATCC 161
Qy	268 AGGTGGCAAGAGGTGATGAAGCCACAGAGCCGCTCCTCAACCTCAACGTACAC 327
Db	162 TGTGGCAAAGAGGTGATGAACGACCTTCTGGAGCTGCTGGTATACCA 221
Qy	328 TCAAGTCCAGTCCAGACAGCTGTCACCTTCTGCGGGCTCTCCACCTCAAGTGGCC 387
Db	222 TCAAGTCCAGGCCAGAACGCTTCTGGTACAGGAACTCTGGTACCT 281
Qy	388 ATGGGACAATGGCAGGTACAGTGTGGAGCTGTGGTCAAGCCAGTGTCAGC 447
Db	282 ATGGACCCAGCAGGTTCCAGATGTACAGGAACACTGGGCTAACCCAGTGTCTAGC 341
Qy	448 TGGGGCCAACCTCACTCTGGAGGAGGAGGGCAGGGCCAGGGCAAGGATGATCTGC 507
Db	342 TGGGTGTACTCTGCTACGCTACGGTGGCTGAGCTGGCTGCTGC 401
Qy	508 AGGGTCTGGAGAGCCACCTATACCAACAGCCCTGATCGGAAGGATGGGAGGTC 567
Db	402 TGGCATCTCAGSCAGGCCCTACCTACCTACGGCTGGTGGGAATGGTGGCTTC 461
Qy	568 ACCTGCAAGAGAACCTATGCCAGGGCCCTGCCAACTCTCTGGCAGGCGAGA 627
Db	462 TGGCACGAAAAGCCACTTCATGGAAAAACGCCAACTCTCCCTGGCTGCCAGA 521
Qy	628 CATGG 633
Db	522 CCTCTG 527

Query Match 26.3%; Score 261.2%; DB 14; Length 384;
 Best Local Similarity 98.6%; Pred. No. 1.e-27;
 Matches 273; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Source	FEATURES
Qy	693 GTGCCCTGAGTGGTGA-CCTGAGAGTGGGACTCTGGAGGGCTCCCTGGAGGAGGCCAT 751
Db	278 GTGGCCCTAGTGGTACCCGAAGATGGAGCTGGTCCCTGGAGGCCAT 219
Qy	752 CCTTGCTTGGCTGCTGAGGACTTGGGAGGAGTTGGGGGTT 811
Db	218 CCTTGCTTGGCTGCTGAGGACTTGGGAGGAGTTGGGGGTT 159
Qy	812 CAGGATAGGGATGGGAGGACTTGGGAGGCTGAGTGAACCGTC 871
Db	158 CAGGATAGGGATGGGAGGACTTGGGAGGCTGAGTGAACCGTC 99
Qy	872 CAGAGGCCAAGGCCAGGGACTTGGGAGGCTGAGTGAACCGTC 931

Db	98	CAGAGAGCCAGGCAGGGACTGAGGCCATAGGGCACTGTTGTATTGGAG	39		Db	200	CTTGGCTTGCCTCTACAGGAGCACCGCCCTGAGTGAAGGAGTTGGGGTTC	141			
Qy	932	TTCATGCAAAATGAGTGTGTTAAGCTGCTTGTGCA	968		Qy	813	AGGATGGGAAATGGGAGTCAGGGACCAAGCAGCCATGAGTGAACGCC	872			
Db	38	TTCATGCAAAATGAGTGTGTTTAGCTGCTTGTGCA	2		Db	140	AGGATGGAAATGGGAGTCAGGGACCAAGCAGCCATGAGTGAACGCC	81			
RESULT 16					Qy	873	AGAGAGCCAAGCACGGAGACTGGAGGCCATGAGCTGACTGTTGAGT	932			
H5B715/C	H5B715	284 bp mRNA	linear EST 05-OCT-1995		Db	80	AGAGAGCCAAGCACGGAGACTGGAGGCCATGAGCTGACTGTTGAGT	21			
LOCUS	Y25C07.s1	soares fetal liver spleen INFSL Homo sapiens cDNA clone IMAGE:2063116 3 , mRNA sequence.			Qy	933	TCACTGCAAATATGGTGTGT	952			
DEFINITION					Db	20	TCACTGCAAATATGGTGTGT	1			
ACCESSION	H5B715										
VERSION	H5B715.1	GT:1011547									
KEYWORDS	EST.										
SOURCE	Homo sapiens (human)										
ORGANISM	Homo sapiens										
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	(bases 1 to 376)				RESULT 17						
REFERENCE					T80187/c	T80187	376 bp mRNA	EST 15-MAR-1995			
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marras,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.				LOCUS	Ydb8503.s1	Soares fetal liver spleen INFSL Homo sapiens cDNA clone IMAGE:115060 3 , mRNA sequence.				
TITLE	The WashU-Merck EST Project				DEFINITION						
JOURNAL					ACCESSION	T80187					
COMMENT	Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 912 High quality sequence stops: 241 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 912 Std Error: 0.00 Seq primer: Promega -21m13 High quality sequence stop: 912 High quality sequence stops: 241 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 912 Std Error: 0.00 Seq primer: Promega -21m13 High quality sequence stop: 912 High quality sequence stops: 241 Location/Qualifiers 1. .284 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:3775447" /db_xref="taxon:9606" /clones="IMAGE:206316" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /clone_lib="Soares fetal liver spleen INFSL" /note="Organ: Liver and Spleen; Vector: pRT73D (Pharmacia) with a modified polylinker; Site: 1: Pac I; Site: 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGAAAGAATTAAATAATTCATCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo." BASE COUNT ORIGIN										
FEATURES source					source	1.	.376 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:470677" /db_xref="taxon:9606" /clone="IMAGE:115060" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /clone_lib="Soares fetal liver spleen INFSL" /note="Organ: Liver and Spleen; Vector: pRT73D (Pharmacia) with a modified polylinker; Site: 1: Pac I; Site: 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACAGAAAGAATTAAATAATTCATCTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo." BASE COUNT ORIGIN				
FEATURES source					source	2.	.376 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:470677" /db_xref="taxon:9606" /clone="IMAGE:115060" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /clone_lib="Soares fetal liver spleen INFSL" /note="Organ: Liver and Spleen; Vector: pRT73D (Pharmacia) with a modified polylinker; Site: 1: Pac I; Site: 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACAGAAAGAATTAAATAATTCATCTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo." BASE COUNT ORIGIN				
Query Match	25.9%	Score 256.9; DB 14; Length 284;			Query Match	23.7%	Score 235.2; DB 14; Length 376;				
Best Local Similarity	99.2%	Pred. No. 8.7e-27;			Best Local Similarity	94.4%	Pred. No. 8.2e-24;				
Matches	258;	Conservative 0; Mismatches 2; Indels 0; Gaps 0;			Matches	0; Mismatches 14; Indels 1; Gaps 1;					
Qy	693	GTCGCCCAAGGTGGTACCAAGAGATGGAGCTGGCGCTCTGAGTGAAGGAGTTGGGGTTC	752		Qy	694	GTCGCCCAAGGTGGTACCAAGAGATGGAGCTGGCGCTCTGAGTGAAGGAGTTGGGGTTC	201			
Db	260	GTGGGCCAGGGTGGTACCAAGAGATGGAGCTGGCGCTCTGAGTGAAGGAGTTGGGGTTC	812		Db	261	GTGGGCCAGGGTGGTACCAAGAGATGGAGCTGGCGCTCTGAGTGAAGGAGTTGGGGTTC	201			
Qy	753	CTTGCCCTGCCTCTACAGAGCACCGCCGCTCTGAGTGAAGGAGTTGGGGTTC	812		Qy	754	CTTGCCCTGCCTCTACAGAGCACCGCCGCTCTGAGTGAAGGAGTTGGGGTTC	812			

/sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /clone_lib="Soares mouse embryo NbME13.5 14.5"
 /note="Vector: pMT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTGCAATCTGAATGGAGCGGCCAAATTTCCTTTTTTTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne State Univ., from 2 13.5dpc double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pMT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "
 185 a 251 c 470 g 208 t 47 others
 BASE COUNT ORIGIN
 Query Match Best Local Similarity 22.0%; Score 218.4; DB 14; Length 1161;
 Matches 330; Conservative 0; Mismatches 131; Indels 13; Gaps 2;
 QY 66 GGGCTCCTGGCTCTGGCTGCTGGCTGGCTCCAGGGCTCTCAGGGCTGG 125
 Db 6 GGGCTCCTTCAGCTGCTCTGGCTCATGGCTGGCTCCAGGGCTGG 65
 QY 126 GAGGAAATAATACCCCTGTTGTTCTCAAGTCATAAAGTCCTGGAAAGTTCCCCAA 185
 Db 66 ACAGAGGCATCA-----CTATGGCTACAAGTACTGGAGTTATCCCAA 113
 QY 186 GGCGCTGGTGGCTCATAACTGGCTGGCAACCCAGCACACCCGCTATCCCTATCC 245
 Db 114 AGCCG-GAGGTGCTTATAACCTGGCTGGCTGGCTCCAGGCCATCATCT 172
 QY 246 CTCCTGGAAACAAAGAACATAGGTGGCCAAGAGTGGTGAAGACCCAGGCC 305
 Db 173 CTCTCTGGCTAGCGGAGSTATCTGGTGGCAAAAAGTTGGCATGACTCGTGC 232
 QY 306 TCCCTTAACCTCAACCTCAACACTCAAGTCAGCTCCAGCTCCACCTTCCTCGGG 365
 Db 233 TCCCTTAACATCAATATCACATCAAGTCAGCTCCACCTCACTTCCTCGCAG 292
 QY 366 GCGTCTCCACCTCAAGTCAGTGCCCATGTGGCATGGCTGGCTGGCTG 425
 Db 293 GCAACCTCGAACCTCGACACTCGCACCTATGGACACCTGGCTCCAGTAGGAACTG 352
 QY 426 TGGTCCAAGCCAGTGTGAGTCAGTCTGGCAGGCGCAACTTCAGTCAGGGAGGC 485
 Db 353 TGGGTTAAGCCAGTGTCTCACTGGGGTCACTTCAGGACTCGGG 412
 QY 486 CCCAGCTGGAGATGATGCTGGCAGGCGTCTGGCAAGCCACCPATCACCAC 539
 Db 413 CGCACTGTGGAGCTCTGGCATCCCTAGGAGCCCATCATCACC 466
 QY 22 AA050499 RESULT 22
 LOCUS AA050499
 DEFINITION mu20f02_r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
 ACCESION AAB50499
 VERSION AAB50499_1 GI:1530170
 KEYWORDS EST,
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Scurognathii; Muridae; Murinae; Mus.
 REFERENCE Schellekens, K.; Kucaba, L.; Allen, M.; Bowles, M.; Dietrich, N.; Dubuque, T.; Geissel, S.; Eustache, J.; Inacy, M.; Le, M.; Martini, J.; Morris, M.; Scheelekens, B.; Steptoe, M.; Tan, F.; Underwood, K.; Moore, B.;
 AUTHORS 448 TGCGGGCCAAACTTCATCTCTGGAGACAGAGGGCAAGCCAGGGTGGAGATGTCGCC 507
 340 TGAGGGTGAATCTGGCTCTACGGCATGGGACTCTGGGCCACTGGCTCTGCC 399
 508 AG 509

Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibara,K., Shirakai,T., Tagami,M., Waki,K., Watanuki,A., Muramatsu,M. and Hayashizaki,Y.	Submission	Direct
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)		
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)		
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)		
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-29 (2001)		
cDNA was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.		
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge), whose assistance we gratefully acknowledge.		
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.		
Location/Qualifiers		
1. .364		
/organism="Mus musculus"		
/mol_type="mRNA"		
/strain="NOz"		
/db_xref="Taxon:10090"		
/clone="FB30016F14"		
/tissue_type="activated spleen"		
/clone_lib="RIKEN full-length enriched, activated spleen"		
84 a 123 c 82 g 74 t 1 others		
FEATURES		
source		
BASE COUNT		
ORIGIN		
Query Match		
Best Local Similarity		
Matches 256; Conservative 0; Mismatches 100; Indels 12; Gaps 1;		
Qy 44 GCCTAACCCACCGAGCAAGGCTCTGGCTGTCTGGCTGGCTGGCTGC 103		
Db 9 GCCCCAACCTCACAGACGCCATGGGCTCTTCAGTGTCTTCATGGCTGC 68		
Qy 104 CAGGAGTTCTCAGGGACGGAGGAATAACCCCTGTTCTCATTCCTACAA 163		
Db 69 CTGGCTCTCAGAGGAGAGCATA-----CCATGGCTACAA 116		
Qy 164 AGTCTCTGGAAATTTCCCACAGGCCGCTGCTATACTCTGCTGTGACCCAGCC 223		
Db 117 AGTACTGAAATGTTATCCCCAACGCCGAGGGCTTATACTCTGCTGTGAGGC 176		
Qy 224 ACCACGGCCATGACCTATTCCCTCTGGAAACAGACATAAGGTGCCAGAGGT 283		
Db 177 GTGCCAGGCCATCACATCTCTCTCTGGCTAGCCGAGTATCTGGTGGCAAAGGT 236		
Qy 284 GTGAAAGCCACGAGCCGGCCTTCAACCTCAACGCTCACCTCAAGTCCAGA 343		
Db 237 TGTGATGACTCCGCTGCCACCTCTTCACATCAAGTCCAGCAGA 296		
Qy 344 CCCTGCTCACCTACTCTGCGGGCTCTCACCTCAAGTCCATGTGGACAGTGGCCAG 403		
Db 297 CCCTGCTCACCTACTCTGCGGGAAACCTCGAACCTATGACCCAGCAGNAG 356		
Qy 404 GCTACAGA 411		
Db 357 GGTCCAGA 364		
Query Match	15.5%	Score 154; DB 12;
Best Local Similarity	75.2%	Length 537;
Matches 206; Conservative 0; Mismatches 65; Indels 3; Gaps 1;		
Qy 433 AGCGAGTGTCTGAGCTGGCGGCAACTTCACCTGAGACAGGG--GGCAAGCCCA 489		
Db 132 AGCGCGTGTCTCCAGCTGTGTGTTAACCTTCATCTGCTGAGCGGGCTCT 191		
Qy 490 GGTTGAGATGATGTCGCGGCCACCTATCACACAGCAGCTGATGCTG 549		
Db 192 GGTTGAAGTGTGTGGATCTGGCACTTCACCCATCACCTAACGTTGTCG 251		
Qy 550 GGAAAGGATGGGAGGTCCACCTGAGGAGCCATGCCACAGGAGCCTGCGCAACTCT 609		
Db 252 GGGGGAGATGGCCACCTACATGAAGAGACATACAGAAGTCACTACAGGGAGCTGCTG 311		
Qy 610 CCTTCCTGCGGCCACCTATCACACAGCTGGCTGCTGCTGCTGCTGCTG 549		
Db 312 CCTTCCTGCTCACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 371		
Qy 670 TCCAGCACAGGCCCTACAGTGCTGCCCCAGG 703		
Db 372 CCCAGCACAGTGCCCTACACCTGCTGCCCCAGG 405		

SOURCE	Homo sapiens (human)	TITLE	cDNAs from fetal heart (1996)
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	COMMENT	Unpublished Contact: Liwei CC Brigham and Women's Hospital Harvard Medical School 75 Francis St. Boston, MA 02115, USA Tel: 617/328915 Fax: 617/50935 Email: cliwei@rics.bwh.harvard.edu
REFERENCE	1 (bases 1 to 540)	FEATURES	Seq primer: GAAATTAACCCNACPAAGGG.
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/cnicgap.	source	Location/Qualifiers
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		1. .194 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606!" /clone="K3475" /lab_host="E. coli XL1-Blue"
JOURNAL	Unpublished		/note="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2: XbaI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XbaI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XbaI, for directional cloning into pDigest Lambda ZAP Express."
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaps-@mail.nih.gov	BASE COUNT	51 a 43 c 63 g 37 t
CDNA Library Preparation:	The I.M.A.G.E. Consortium/LINL DNA Sequencing by National Institutes of Health Intramural Sequencing Center (NISC)	ORIGIN	
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LINL at: found through the I.M.A.G.E. Consortium/LINL at: info@image.lnl.gov	Query Match 14.6%; Score 145.2; DB 14; Length 194;	Best Local Similarity 96.5%; Pred. No. 4.6e-11;	
Seq. Primer: -21M13 forward primer (ABI).	Matches 191; Conservative 0; Mismatches 3; Indels 4; Gaps 4;	Matches 191; Conservative 0; Mismatches 3; Indels 4; Gaps 4;	
Location/Qualifiers	1. .540 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606!" /clone="IMAGE:3289604" /sex="male" /dev_stage="adult" /lab_host="DHL0B" /clone_lib="NCI CGAP Pr28" /note="Organ: prostate; vector: pTT3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP P-22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clone IDs 985608-986759, 1101192-1101959, and 1217928-1220515); clones made from the same library (clone IDs 985608-986759, 1101192-1101959, and 1217928-1220515); subtraction by Bento Soares and M. Fatima Bonaldo."	Qy 750 ATCCCTGGCCTTGCGCTTACAGGAGCACCGCCCTGAGTGAAGAAGGAGTTGGGG 809 Db 1 ATCCCTGGCCTTGCGCTTACAGGAGCACCGCCCTGAGTGAAGAAGGAGTTGGGG 59	
FEATURES	BASE COUNT 126 a 113 c 128 g 173 t	Qy 810 TTCAAGATTAGGAATGGGGAGTCAGAGGAGGCAANGCAGGCCATGTAGAATGACCG 869 Db 60 TTCAAGATTAGGAATGGGGAGTCAGAGGAGGCAANGCAGGCCATGTAGAATGACCG 119	
source	Query Match 15.5%; Score 153.8; DB 14; Length 540;	Qy 870 TCCAGAGGAGGCCAGAGGAGGACTCAGGCCATCAGGCTGAGTCGTTGATTTGG 929 Db 120 T-CAGAGGAGGCCAGAGGAGGACTCAGGCCATCAGGCTGAGTCGTTGATTTGG 176	
	Best Local Similarity 98.7%; Pred. No. 1.8e-12;	Qy 930 AGTCATGCAAATGAGT 947 Db 177 AGTCATGCAAATGAGT 194	
RESULTS	Matches 155; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	RESULT 32 BB842514 LOCUS BB842514 DEFINITION RIKEN full-length enriched, 6 days neonate spleen Mus DB BB842514 DEFINITION musculus cDNA clone F430005P21 5', mRNA sequence. ACCESSION BB842514 VERSION BB842514.1 KEYWORDS GL:17043245 EST. SOURCE Mus musculus (house mouse) ORGANISM Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanasaki,T., Ishii, Hayatsu,N., Hiramoto,K., Hirokawa,T., Hirorozane,T., Imotani,K., Kojima,Y., Kawai,J., Konno,H., Kondo,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibusawa,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanuki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y., RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)	
REFERENCE	1 (bases 1 to 194)	JOURNAL	Unpublished
AUTHORS	Liwei,C.	COMMENT	Contact: Yoshihide Hayashizaki

Locus	BB843270	370 bp	mRNA linear	EST 26-NOV-2001
Definition	BB843270 RIKEN full-length enriched, 6 days neonate spleen Mus musculus cDNA clone F430010H21 5'			
Accession	BB843270			
Version	BB843270.1			
Keywords	EST.			
Source	Mus musculus (house mouse)			
Organism	Mus musculus			
Reference	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 370)			
Authors	Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirakawa,T., Ito,M., Itoh,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuura,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaka,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sobue,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akabira,S., Tanaka,T., Tomaru,A., Toya,T., Watanuki,A., Yasunishi,A., Muramatsu,M., and Hayashizaki,Y.			
Title	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1777-1777. (2000)			
Journal	Unpublished			
Comment	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) Tel: 81-45-503-9222 Fax: 81-45-503-9216			
Features	Location/Qualifiers			
	1 . 382 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="F430005P21" /tissue_type="spleen" /dev_stage="6 days neonate" /clone_lib="RIKEN full-length enriched, 6 days neonate spleen" e mouse tissues.			
	87 a 122 c 99 g 74 t			
	Source			
	Query Match 14.4%; Score 143; DB 10; Length 382; Best Local Similarity 69.8%; Pred. No. 7e-11; Gaps 1; Matches 213; Conservative 0; Mismatches 80; Indels 12; Gaps 1;			
Dy	44 GCCCTAACCCACGGGATGGGGATGGGGCTCCCTGGGTGTTCTGCTTGGCCCTGGCTGCTGC 103 90 GCCCCAAACCTCACAGCATGGCTCCCTCAAGTGCTCTTCAGTGCTGCTCTGGCTGCTGC 149			
Dy	104 CAGGAGCTCTCCAGGCCGGGGAAATAAACCTGTGGTCTCATGTCTACAA 163 150 CTGGCTCTCACAGGAGAACAGAGCTCA 197			
Dy	164 AGTCTGGAGTTTCCCAAAGGCCGCTGGCTCATACCTCTGTGCTGACCCAGCC 223 198 AGTACTGGAGTTATCCAAAGCGGAGGGTATAACCTGATGCCCTGAGG 257			
Dy	224 ACCACCGCCATCACCTATTCCTGTAACGAACTCAAGTGGCCAAAGGT 283 258 GTCCAGCCATCACATCTCTCTGGAGCTTACCTGGTCAACATCCAGCAGCCAGA 317			
Dy	284 GTGAGAGCCAGAGCCGCTGGCCCTCTGGTCAACATCCAGCAGCCAGA 343 318 TGTGATGACTCGTGGCCCTCTGGTCAACATCCAGCAGCCAGA 377			
Dy	344 CCTGCG 348 Db 378 CCTGCG 382			
Query Match 13.4%; Score 132.6; DB 10; Length 370; Best Local Similarity 68.9%; Pred. No. 2e-09; Gaps 1; Matches 202; Conservative 0; Mismatches 79; Indels 12; Gaps 1;				
Qy	44 GCCCTAACCCACGGGATGGCTCCCTGGCTGCTGCTGC 103 90 GCCCCAAACCTCACAGCATGGCTCCCTCAAGTGCTCTGGCTGCTGC 149			

Qy	104	CAGGAGCTTCCAAAGGCCACGGGAGGAATAATTACCCCTGTGGTCCTCAATGGCTTACA	163	Qy	485	CCCCAGGGTGGAGATGATCTGCCAGGGTCTGGGAGGCCACCTATCACCACAGCT	544
Db	150	CTGGGCTTCAGAGGGAGACAGGGATCA-----CCATGGCTTACA	197	Db	283	CCCTTGCTGAGGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	342
Qy	164	AGTCTCTGGAAAGTTTCCCAGGCCGCTGCTCATAACTGCTGCTGCTGCTGCTGCTG	223	Qy	545	GATCGGGAGGATGGGAGGTCACTGAGGAGGATGCCAGGGAGGATGCCAGGGAGG	604
Db	198	AGTACTGAAAGTTTCCCAGGCCGCTGCTCATAACTGCTGCTGCTGCTGCTGCTG	257	Db	343	TGTCGGAGGAGGAGCTGCGACTACATGAGGAGGAGGAGGAGGAGGAGGAGG	402
Qy	224	ACCAACGCCATACCTATTCTCTGTAAGTGGCAAGAACATAAGTGGCAAGAGGT	283	Qy	605	CTTCTCTTCTCTGC	618
Db	258	GTCAGGCGCTACATACCTCTCTCTGCTGCTGAGGTATCTGTGGAAAGAGGT	317	Db	403	CTTCTCTTCCCGC	416
Qy	284	GGTGAAGAACCAAGGCCGCTCTTCAACCTCAACGTCACACTGAGTC	336				
Db	318	TGTCATGACTCCGTGGCCCTTCAACATCAATACCATAGTC	370				
		RESULT 35					
		BF442092					
		LOCUS	420 bp	mRNA	linear	EST 01-DEC-2000	
		DEFINITION	BF443482	MARC 2PIG Sus scrofa cDNA 5'	mRNA sequence.		
		VERSION	260980	MARC 2PIG Sus scrofa			
		KEYWORDS	BF443482	EST.			
		ORGANISM	BF443482.1	GI:11503574			
		REFERENCE	1	(bases 1 to 420)			
		AUTHORS	Fahrenkrug,S.C.,Smith,T.P.L.,Frekking,B.A.,Cho,J.J.,White,J.J.,Vallet,J.,Wise,T.,Rohrer,G.A.,Pertea,G.,Sultana,R.,Quackerbush,J.,and Keele,J.W.				
		TITLE	Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly				
		JOURNAL	Genome 13 (8), 475-478				
		MEDLINE	22213789				
		PMID	12226715				
		COMMENT	Contact: Smith TPL				
			USDA, ARS, US Meat Animal Research Center				
			PO Box 166, Clay Center, NE 68933-0166, USA				
			Tel: 402 762 4366				
			Fax: 402 762 4390				
			Email: smith@email.marc.usda.gov				
			Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.				
			PCR PRIMERS				
			FORWARD: AGAAAACAGCTATGACCAT				
			BACKWARD: GTTTTCCCGTACGAGC				
			Plate: 93 row: K column: 24				
			Seq primer: ATTAGGTGACATATAG.				
			Location/Qualifiers				
			1..503				
			/organism="Sus scrofa"				
			/mol_type="mRNA"				
			/db_xref="taxon:9823"				
			/tissue_type="pooled"				
			/lab_host="DEI0B"				
			/clone_lib="MARC 2PIG"				
			/note="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue From testis, ovary, endometrium, hypothalamus, pituitary, and placenta."				
			BASE COUNT	104 a	160 g	79 t	
			ORIGIN				
			Query Match	9.3%	Score 92.6;	DB 10;	Length 503;
			Best Local Similarity	70.8%	Pred. No. 0.00072;		
			Matches	138	Mismatches	54	
			FEATURES				
		source	Location/Qualifiers				
			1..420				
			/organism="Sus scrofa"				
			/mol_type="mRNA"				
			/db_xref="taxon:9823"				
			/tissue_type="pooled"				
			/lab_host="DRI0B"				
			/clone_lib="MARC 2PIG"				
			/note="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue From testis, ovary, endometrium, hypothalamus, pituitary, and placenta."				
			BASE COUNT	99 a	109 c	130 g	82 t
			ORIGIN				
			Query Match	9.4%	Score 93.2;	DB 10;	Length 420;
			Best Local Similarity	71.1%	Pred. No. 0.00054;		
			Matches	138	Mismatches	53	
			FEATURES				
		source	Location/Qualifiers				
			0..53				
			/organism="Sus scrofa"				
			/mol_type="mRNA"				
			/db_xref="taxon:9823"				
			/tissue_type="pooled"				
			/lab_host="DRI0B"				
			/clone_lib="MARC 2PIG"				
			/note="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue From testis, ovary, endometrium, hypothalamus, pituitary, and placenta."				
			BASE COUNT	428	GGTCAANGCCAGTGTGCTGAGCTGGCTGGGGCCAACTTCACTCTGGAGGAGG--GGCAGG	484	
			ORIGIN	223	GCCAGACCCCGTCCAGCTGTGCTTAACTTCACTCTGGACGCCCTGGG	282	
			Query Match	9.4%	Score 93.2;	DB 10;	Length 503;
			Best Local Similarity	70.8%	Pred. No. 0.00072;		
			Matches	138	Mismatches	54	
			FEATURES				
		source	Location/Qualifiers				
			0..53				
			/organism="Sus scrofa"				
			/mol_type="mRNA"				
			/db_xref="taxon:9823"				
			/tissue_type="pooled"				
			/lab_host="DRI0B"				
			/clone_lib="MARC 2PIG"				
			/note="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue From testis, ovary, endometrium, hypothalamus, pituitary, and placenta."				
			BASE COUNT	369	GGGTGAAGGTGTCGTGCTGGCATCATCACGTCAGTGTGCTG	428	
			ORIGIN				
			Query Match	9.3%	Score 92.6;	DB 10;	Length 503;
			Best Local Similarity	70.8%	Pred. No. 0.00072;		
			Matches	138	Mismatches	54	
			FEATURES				
		source	Location/Qualifiers				
			0..53				
			/organism="Sus scrofa"				
			/mol_type="mRNA"				
			/db_xref="taxon:9823"				
			/tissue_type="pooled"				
			/lab_host="DRI0B"				
			/clone_lib="MARC 2PIG"				
			/note="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue From testis, ovary, endometrium, hypothalamus, pituitary, and placenta."				
			BASE COUNT	490	GGGTGAAGGTGTCGTGCTGGCATCATCACGTCAGTGTGCTG	549	
			ORIGIN				

Qy 550 GGAGGATGGGAGGTCCACCTGAGGAGCATGCCACAGGACCCGCCAACTTC 609
 Db 429 GAGGGGACATGGCACCTCATGAAGCAGACAGAACTCAAGAGGTCGCAACTCT 488
 Qy 610 CCTTCCCTGCGAACCT 624
 Db 489 CCTTCCCGCTCACCC 503

RESULT 36
 AA031034 AA031034 208 bp mRNA linear EST 21-AUG-1996
 DEFINITION mi47h05_r1 Soares mouse embryo NM_0013.5 14.5 Mus musculus cDNA
 ACCESSION AA031034.1
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 208)
 REFERENCE Marr,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Martin,J., Morris,M.,
 Scheellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Thaising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

AUTHORS Unpublished
 TITLE WashU-HMM Mouse EST Project
 JOURNAL Contact: Marra M/Mouse EST Project
 COMMENT WashU-HMM Mouse EST Project
 Washington University School of Medicine P
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LiNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGJ: 280529

Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 193.

FEATURES source
 /note="Soares mouse embryo NBME13.5 14.5"
 /note="vector: pTT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5,
 TGTACCAACTGAGTGCGAGCGGCCGAAATTTCCTTTTTTTTT
 T 3], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Mirnor Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT773 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fatima Bonaldo."

BASE COUNT 47 a 61 c 51 g 49 t
 ORIGIN

Qy 137 TACCCCTGTTGCTCCATTGCTTACAAAGTCTGGAGTTTCCCCAAAGGCCGCTGGT 196
 66 ----- -TACATTGCTAACAGTACTGAAAGTTATCCCAAAGCGGAGGTA 114
 197 GCTCATAACCTGCTGTGCAACCCASGCCACCCGCATCACATTCCCTCTGGAAC 256
 115 GCTTAAACCTGCAAGGCCCCTGAGCCATCATCATCTCTGGCT 174

Qy 257 CAAGACATCAAGGTGGCCAGAAGTGGTGA 289
 Db 175 CGGAGGTATCTGGGGCAAAAATGGTGGCA 207

RESULT 37
 AI187208/c LOCUS AI187208 483 bp mRNA linear EST 10-NOV-1998
 DEFINITION qf28b11.x1 Soares testis NTR Homo sapiens cDNA clone IMAGE:1751325
 3', similar to contains MER27.t1 MER32 repetitive element ; mENA
 sequence.
 AI187208 1 GI:3737846
 ACCESSION AI187208
 VERSION EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Bacteria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 483)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/hcicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LiNL at:
 www-bio.lnl.gov/bbrp/image/image.html
 Insert Length: 678 std Error: 0.00
 Seq primer: -40UP from Gibson
 High quality sequence stop: 381.
 FEATURES source
 1. organism="Homo sapiens"
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /clone="Soares testis_NTR"
 /clone_lib="Soares testis_NTR"
 /clone_host="DH10B"
 /note="vector: pTT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 Inc., and primed with a Not I - oligo(dT) primer [5,
 TGTACCAACTGAGTGCGAGCGGCCAATTTTTTTTTTT
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT773 vector. Library
 went through one round of normalization to Cots, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 133 a 103 c 78 g 169 t
 ORIGIN

Query Match 8.0%; Score 79.8; DB 9; Length 208;
 Best Local Similarity 65.3%; Pred. No. 0.06%;
 Matches 139; Conservative 0; Mismatches 62; Indels 12; Gaps 1;

Qy 829 AGGTGAGGAGCAAGGACGGGAGCTCTCAAGGCACGGAGGAAAT 136

Query Match 5.8%; Score 57.8; DB 9; Length 483;
 Best Local Similarity 67.9%; Pred. No. 56;
 Matches 112; Conservative 0; Mismatches 47; Indels 6; Gaps 2;

Qy 77 GCTGTCTGCTGGCGCTGCTGGCTGCCAGAGCTCTCAAGGCACGGAGGAAAT 136

Db	284	AGGTAGACATAGAAAGTAGCAGCTATAGGATGAACTAGGGTAAACTCAGGGATGTAAGCAGACATAAGTGAAT 225	Db	461	GGTGGGAAAGAATGGGAGGAAACTCAGGGATGTAAGCAGACATAAGTGAAT 402
Qy	887	GGCAGGAGCTGCAAGGCCATTAGCGGTGACTGTTGGAGTTATGMAAATGAG 946	Qy	864	GAACCGTCCAGAGCCAGCAAGGCAAGGATGCAAGGCCATGGAGCTGCAGCTGTCG 923
Db	224	ACATGAGGACTGTAAGGTAATTAATGTACT---GATGTTGGAGTTCATGCTTAATGAG 169	Db	401	GATAAAAGCTCTAGTTTCTATGCAATGAGACTACAGATAATAATATGCTGT-GT 343
Qy	947	TGTGTTTAGCTGCTCTGCACAAAAAAAGAAAAAAA 991	Qy	924	ATTTGGAGTTCTGCAAAATGAGTGTTTACCTGCTCTTGCACAAAAAAAGAAAAA 983
Db	168	TAGATTTAGCTGCTCTGCACAGAACAAAATCTAAGGGAA 124	Db	342	ATTTGGATTCTGATATAATGAGTAGGTTTATGCTGTCAGCACAAAGCTAAATAA 283
			Qy	984	AAAAAAA 991
			Db	282	AAACAAA 275
RESULT	38				
AWT94912;c					
LOCUS		AWT94912 466 bp mRNA linear EST 16 MAY 2000			
DEFINITION		RC6-JM0015-210200-011-D07 UM0015 Homo sapiens cDNA, mRNA sequence.			
ACCESSION		AWT94912			
VERSION		AWT94912.1			
KEYWORDS		GI:7846691			
SOURCE		Homo sapiens (human)			
ORGANISM		Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE		1 (bases 1 to 466)			
AUTHORS		Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.			
TITLE		Shotgun sequencing of the human transcriptome with ORF expressed sequence tags			
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)			
MEDLINE		10202663			
PUBLMED		10737800			
COMMENT		Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, São Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br			
		This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC6-UM0015-210 20-011-D07&t3=2000-02-21&c4=1) Seq primer: puc 18 forward High quality sequence start: 14 High quality sequence stop: 462. Location/Qualifiers			
FEATURES		1. .466 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /dev_stage="Adult" /clone_lib="UM0015" /note="Organ: uterus; Vector: puci18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the PUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
BASE COUNT		113 a 98 c 171 t			
ORIGIN		5.3%; Score 52.8; DB 9; Length 466;			
		Query Match 5.3%; Score 52.6; DB 2; Length 493;			
		Best Local Similarity 58.5%; Pred. No. 2.4e+02;			
		Matches 110; Conservative 0; Mismatches 77; Indels 1; Gaps 1;			
Qy	804	GGGGGGTAGGATGGGAACTAGGAGCTAGGACCATGTAAT 863	Qy	820	GGAATGGGGAGGTCAGGAACTAGGAGCTAGGACCATGTAATGACCCGTCAGAGAGC 879

Db	174	GGGAGATGTAGGGTCGAGGATACAAAGTACAGATATGAGATGAAACAGTC--TACA	118	Qy	873	AGAGAGCCAGCAGGCAAGGACTGCAGGACTGCGACATGCCATAGGCCTGCACTGTTGTTGAGT	932
Qy	880	CAAGCACGGAGGACTGCAGGCCATACGGCTGCACTGTTCGTATTGGAGGTCATGCA	939	Db	480	AGAG-GCCATAATGTCATGAGGACTCTAGTCACACATTGA--ATTATACGGGGAT	425
Db	117	GAATATGTCATGAGGCTCTAGGAACTAACATTGATGTTGAAATCATGCT	58	Qy	933	TCAATGCAAATGAGTTGTTAGCTGTTAGCTGCACTGCTCTGCATATTGAA	992
Qy	940	AAATGAGTGTGTTAGCTGCTCTGCCATACAAAAAA	982	Db	424	TTCGCTTAGACAGTAAATGAGTTAGAGCTCTGCATATTGAA	365
Db	57	AAAGGACTAACATTAGCTGCTCTGCCAAACAAAGAAA	15				
					RESULT 41		
					BY560588		
				LOCUS	BY560588	443 bp mRNA linear EST 15-DEC-2002	
				DEFINITION	BY560588	RIKEN full-length enriched, activated spleen Mus musculus cDNA clone F83028K04 3', mRNA sequence.	
				ACCESSION	BY560588		
				VERSION	BY560588.1	GI:26895770	
				KEYWORDS	EST.		
				SOURCE	Mus musculus (house mouse)		
				ORGANISM	Mus musculus		
				REFERENCE	Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.		
				AUTHORS	(bases 1 to 443)		
					Okazaki, T., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Kiyosawa, H.		
					Nikaido, I., Furuno, N., Saito, R., Suzuki, H., Yamamoto, I., Kiyosawa, H.		
					Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chorbia, C., Corbani, I.E., Cousins, S., Dalla, B., Dragani, B., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Garibaldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarrow, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedierski, R.M., King, J.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Leinhardt, B., Lyons, P.A., Magliott, D.R., Maitaisi, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pearce, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C.A., Setou, M., Shimada, K., Sutana, R., Takenaka, Y., Taylor, M.S., Teasdile, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynnshw-Boris, A., Yanagisawa, M., Yang, J., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishiyama, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawaji, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imoto, K., Ishii, Y., Itoh, M., Kagawa, T., Miyazaki, D., Sasaki, K., Sasaki, D., Shibata K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander E.S., Rogers, J., Birney, B., and Hayashizaki, Y.		
				TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
				JOURNAL	Nature 420, 563-573 (2002)		
				PUBLMED	1246683		
				COMMENT	12466831		
					Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) Tel: 81-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Fax: 81-45-503-9222 Email: genome-res@gsc.riken.go.jp/ URL: http://genome-res.gsc.riken.go.jp/		
					Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane , T., Imoto, K., Ishii, Y., Itch, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, K., Nomura, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Wakai, K., Watanuki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission		
					Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)		
BASE COUNT	124	167 c	143 g	172 t			
ORIGIN							
Query Match	5.2%	Score 51.2;	DB 12;	Length 606;			
Best Local Similarity	62.8%	Pred. No. 4.3e+02					
Matches 113; Conservative 0; Mismatches 63; Indels 4; Gaps 2;							
Qy	813	AGGATGGGATGGGAGGTAGAGGCAAGGCAAGGAGCCATGAGATGAACCGGCC	872				
Db	540	AGGAAACGGGAGATGAGGTGAAAGGAGGCAAGGAGCATGATGCC	481				

RIKEN integrated sequence analysis (RISA) system--384-format sequencing Pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000).

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Center and Genome Science Laboratory in RIKEN, a division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

537	CCGCACTTCCTTCTTGCGAGACATGGACTGGTGTGGCCAGCTGCA	656
1	CCAGCCAATTCCTCCGCTTCCAGAACCTGTTCTAGCTGGA	60
657	AAGAAAGCCTAATGTCAGGACAGGCCCTCACAGTGGTGTGCAAGAG	716
61	AACGATGTCGGCTTGACAGCASTGGCCCATCCGTGGCCCAGGCCAG	120
717	ATGGAG	722
121	CTGGTG	126

Genoscope - Centre National de Séquençage :
91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
Web : www.genoscope.cns.fr
The BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
elangaster genome using these BACs. For further information
please see <http://www.trituly.org>. The BDGP Drosophila
elangaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammone in pierre-arnaud Toncini's laboratory at the Drosophila
Genoscope.

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by Partial ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; *cnn bw sp*, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	source	Location/Qualifiers
		/organism="Drosophila melanogaster"
		/mol_type="genomic DNA"
		/db_xref="taxon:7227"
		/clone_id="BACRI9D16"
		/clone_lib="RPCI-98"
		/note="end : TET3"
BASE COUNT	120	a 61 c 61 g 172 t 511 others
ORIGIN		
		Query Match 4.9% Score 49; DB 29; Length 925;
		Best Local Similarity 15.1%; Pred. No. 7.2e+02;
		Matches 47; Conservative 138; Mismatches 127; Indels 0; Gaps 0
Qy	292	CCACAGGCGGGCTTCATCAACCTAACGTACACATCAAGTCCAGACCTGCTCA 351
Db	572	CSCCSSCSSSSCSCBCCCCSSYCCSSSSSSCSSTSCKSVCGTSC 631
Qy	352	CCTACTCTGGGGCTCTCCACCTCACGGTCCCAGTGGACAGTCCAGGCTACAGA 411
Db	632	SSSSSSSSSSSSSSSSSSSTSCKSSSSSSSSSTSKSTSASGSWSAGGGSGS 691
Qy	412	TGCACGGAGGTGTGGTCCAGCCAGTGTCTAGCTGGGCCAACTTCACCTTGAGG 471
Db	692	TGSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSBSCSTSCTSSSSSS 751
Qy	472	ACAGAGGGCAGGCCAGGGTGGAGATGATCTGCAGGGTCTGGCAGGCCACCTA 531
Db	752	SYSSSTSCTCCSYSSSTSSSSSTSWSGTSSTGSSSSSSVGTSSSSDSTSCTCCCTM 811
Qy	532	TCAACCAACAGCTGATCGGGAAAGGATGGCAGGTCACTGAGACCATGCCACA 591
Db	812	CTCCSTSYMBCTSTS CGSSSSSGGGVTKCGGGSSSTNGMBGTSSACSSSSCS 871
Qy	592	GGCGCTGCCA 603
Db	872	SSSSSSSSSS 883

RESULT	43							
CNSN	0052P	CNS0052P	Drosophila melanogaster	844 bp	DNA	Linear	GSS 03-JUN-1999	
LOCUS			DEFINITION	genomic survey sequence TER3		end of BAC #		
			BACLIBRARY of RPCI-98	from Drosophila melanogaster (fruit fly)				
			ALU56652	genomic survey sequence				
ACCESSION		AL006652.1	GI:	4932342				
VERSION			GSS					
KEYWORDS			Drosophila melanogaster (fruit fly)					
SOURCE			Drosophila melanogaster					
ORGANISM			Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriodea; Drosophilidae; Drosophila.					
REFERENCE			1 (bases 1 to 844)					
AUTHORS			Genoscope					
TITLE			Direct Submission					
JOURNAL			Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage - FRANCE					
			BP 191 91006 EVRY cedex - FRANCE					

Determination of this BAC-end sequence was carried out as part of a collaborative project with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information see www.genoscope.cnrs.fr.

melanogaster BAC library was prepared by Kazuyoto Osoegawa and Aaron Mammober in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp; the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

/organism="Drosophila melanogaster"
 /mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone="BACR1P16"

/clone_id="RPCI-98"

/note="end : TET3"

BASE COUNT 261 a 112 c 92 g 35 t 344 others
 ORIGIN 1 - 844

Query Match Score 48 ; DB 29 ; Length 844 ;

Best Local Similarity 17.7% ; Pred. No. 1e+03 ;
 Matches 75 ; Conservative 173 ; Mismatches 167 ; Indels 8 ; Gaps 1 ;

FEATURES source

Qy 534 ACCAACAGCTCTGATCGGAAGATGGCAGGTCCAGTGCAGAAGACATGGCACAGGG 593

Db 359 AASAYSSAASSVAGSSMAAGASCAGCAANMAGCSSSSAARSRMNNSCSCASSS 418

Qy 594 CAGCCCTGCCAACTTCCTCTCGCGGAGCCAGAACATGGACTGTGTGGCAGGGT 653

Db 419 CSVGSSSSAMSSSSCCVVSSCGMASSSSCGGMSAASSSGGSGSRAVERGRNC 478

Qy 654 GCAAACAAGCCAATGTCAGCACAGGCCCTCACAGTGTCGCCAGTAGGTTGACACNG 713

Db 479 CCSMVCCMCSSM-----CCSMCSCSVSSVCAVSSGGSVRSAVGCSTGRTGSSRG 530

Qy 714 AAGATGAGGACTGGTAGGGTCCCCCTGGAGGCCCATCTCTGGCTTGGCGCTAACAGG 773

Db 531 RAGRSGRGCGSSSSVSSGVSSSSSTGMGCACSSVSBSSVASSVSSGBVRGGRRC 590

Qy 774 AGCACCCGCCCTCTGAGTAGAGGAGGTTCAGGATAGGGAAATGGGGAGAGTC 833

Db 591 VGGVGGGSRVSSCSGSSSSSGSGVRSRRSGSANGVRVGSGRRGGGERGGA 650

Qy 834 AGAGGAGCCTAAAGCAGCAGCCATGTAGAAATGAAACCGTCCAGAGGCCAGAG 893

Db 651 NARADDAABRAKARANSHADAKAASSVSSBGRNRRAAVVNGSVSRBNRY 710

Qy 894 GACTGAGGGCATCGCGTGCACATGTTGATTGGAGTCTGATGAAATGAGTGTGTT 953

Db 711 KTKTCKTRKBKTBTCKBKGABWKKCTKKWTRKZKTKTWRGARAADAAGBKKKGKG 770

Qy 954 TAG 956

Db 771 KDG 773

BASE COUNT 180 a 84 c 90 g 117 t

ORIGIN Query Match 4.8% ; Score 47.4 ; DB 9 ; Length 471 ;

Best Local Similarity 59.4% ; Pred. No. 1.6e+03 ; Matches 98 ; Conservative 0 ; Mismatches 66 ; Indels 1 ; Gaps 1 ;

Db 827 GGAGTTCAGGAGCCAAAGCAGCAGCCATGTAGAAATGAAACCGTCCAGAGGCCAGAC 886

Db 35 GAATGTAAGGGATGTAAGGTACAGACATGAATGAAAGTGTAGATTCTTATGT 94

Qy 887 GGCAGGAGGACTGTGAGCCATCAGGTTGACTGTGTTGGATCTGAAATTCAG 946

Db 95 ACACATGAGGACTACAGATAATAATAGTGTCTATGTTGGATTCATGATAAATTCAG 153

Qy 947 TGTGTTTGTGCTGTTGCTGCCAGAAAGAAAAAAAGAAAAAAAGAAAAA 991

Db 154 TAGATTTAGTCGTCAGCCAGCAAGACGAAATACAAA 198

RESULT 45

BX39244/c LOCUS BX39244 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

DEFINITION BX39244 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS011070TM03 3'-PRIME, mRNA sequence.

ACCESSION BX339244 VERSION BX339244.1 KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 471)
 Authors Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Neggi, M. A., da Silva, M. Jr., Zago, M. A., Bordim, S., Costa, F. P., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baita, G. S., Simpson, D. H., Li, W. B., Gruber, C., Jesse, J. and Polayes, D.

TITLE Full-length cDNA Libraries and normalization
 JOURNAL Unpublished

Brunstein, A., de oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M.J., Soares, F., Brentani, R.R., Reit, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prosp. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-27049221
 Fax: +55-11-27070012

Email: asimpson@ludwig.org.br

This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?st1=&c2=RCC6-UM0015-280>)

300-01-C03at3-200-03-2845c4-1)

Seq primer: puc 18 forward

High quality sequence start: 59

High quality sequence stop: 338.

Location/Qualifiers 1. .471
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="UM0015"
 /note="Organ: uterus; Vector: puc18; Site1: Smal; Site2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

stringency conditions."

1. .471

/organism="Homo sapiens"

1. .471

COMMENT	Contact: Genoscope Genoscope - Centre National de Séquençage BP 191 91066 Evry Cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. Contact : Peng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/invitrogen/corporation/1600 Paraday Avenue Genoscope sequence ID : CSODI070AG0ZNP1.	
FEATURES Source	1.. 974 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSODI070TM03" /tissue_type="PLACENTA COT 25-NORMALIZED" /clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized." 69 a 235 c 80 g 310 t 280 others	
BASE COUNT ORIGIN	Query Match Best Local Similarity 4.8%; Score 47.2; DB 13; Length 974; Matches 85; Conservative 85; Mismatches 160; Indels 0; Gaps 0;	
BASE COUNT ORIGIN	Query Match Best Local Similarity 4.8%; Score 47.2; DB 13; Length 974; Matches 85; Conservative 85; Mismatches 160; Indels 0; Gaps 0;	
RESULT 46		
LOCUS	BZ772737 748 bp DNA linear GSS 13-MAR-2003	
DEFINITION	mcv5if02_g1 HFOSMID007 Homo sapiens genomic, genomic survey sequence.	
ACCESSION	BZ772737	
VERSION	GSS. GI:28946421	
KEYWORDS		
SOURCE	Homo sapiens (human)	
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
COMMENT	1 (bases 1 to 748) Cook,J., Delaney,K., Fewell,G., Fulton,L., Magrini,V., Mardis,E., Miner,T., Nash,W., Williams,D. and Wilson,R.K. Unpublished Contact: Richard K. Wilson Genome Sequencing Center Washington University School of Medicine Email: submissions@wustl.edu	
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT	116 a 88 c 93 g 398 t 506 others	

ORIGIN	BASE COUNT	136 a	133 c	128 g	172 t
Query Match Similarity 4.7%; Best Local Similarity 13.1%; Pred. No. 1.6e+13; Matches 69; Conservative 218; Mismatches 237; Indels 1; Gaps 1;	Query Match Similarity 4.6%; Best Local Similarity 55.8%; Pred. No. 2.7e+03; Matches 87; Conservative 0; Mismatches 69; Indels 0; Gaps 0;				
Qy 468 CAGGACAGAGGGCAGCCCCAGGGTGGAGATGATCTGCAGGGCTCGGGCAGCCCA 527	Qy 836 AGGACCCAANGCAAGCCATGATAATGACGCCAGAGGAAAGCGGAGGGA 895				
Db 1080 SVSSSSYAVIVASAVAVAVSVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1021	Db 178 AGAACGGACGGCCGGTGTAGTCAGCTACAGTCCATACAGATGCCAGTCTACAGAGT 119				
Qy 528 CCTATCACCAACAGCCCTGATCGGGAGGATGGCGAGCTCACCTGAGGAGACCATGC 587	Qy 896 CTGCAAGGCCATCAGGGCAACTGTTGTATGGATTCATGCCAAATGAGTGTCTTA 955				
Db 1020 SSANSSVSSSSYAVSSSSAAAASSSSAASSSSAASSSSSSSSSSSSSSSS 961	Db 118 6AAGTGCCCTAGCTTGGAGTGTGTTACTAGAATGTTCAAATAGATTTATA 59				
Qy 588 CACAGGAGCTGCTGCCACTCTCTCTGGAGAGATCATGCAGCTGGTCTGGTC 647	Qy 956 GCTGCTCTTGCACAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAA 991				
Db 960 SAVVSSSSVSSSSAVSSSSAAAASSSSSSSSSSSSVVAASSSS 901	Db 5B AAAACACATGACATAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAA 23				
Qy 648 CAGGTGCAAAACAGCCAAATGTCAGCAAGGGCAAGGGTGGCCCAAGGTGGT 707	RESULT 49				
Db 900 AAAAVASAVAAAAYWSVSTYASSSSVASSVSSSSSANANASSGASATAAWASSSS 841	CD397038/c				
Qy 708 GACCAGAGATGGAGGACTGGAGGTCCCTGGAGGCCCATCTTGGCTTGGCGTC 767	LOCUS CD397038 mRNA linear EST 01-JUN-2003				
Db 840 SAVVSSAA-AVSSMASAASSSSAASSSSSSAGAVSSAKSVASAATSSAGSSAA 782	DEFINITION Gm_ck17110 Soybean induced by Salicylic Acid Glycine max cDNA 3'				
Qy 768 TACAGGAGCACCGCGCGTCAAGTAGAGGAGATTGGGGGTTCAGGATGGAAATGG 827	ACCESSION CD397038				
Db 781 VSSVSTRSSVARTAASSVAAAGSAYASASSASYSSSSMAAAAGSAVSVSSSSAV 722	VERSION GI:31311835				
Qy 828 GAGGTAGAGGAGCAGGCGTAAGGGCGAGCCATGATGAACTGGCCAGGCCAGC 887	KEYWORD EST.				
Db 721 SAAAAASSTSAASASSSSAASSSSSSSSSSSSAASAVVSSSSSSSSSSVRSVASSS 662	ORGANISM Glycine max (soybean)				
Qy 888 GCAGAGACTGAGGCCATGGGTGCACTGTTGGATTGGAGTTCACTCAAATGACT 947	REFERENCE 1 (bases 1 to 626)				
Db 661 SASSSSAAVAAVAAANSANNNSNSNSNSNNNNNAASSSSSANCNNSNNNNNNNN 602	AUTHORS Tian,A.-G., Wang,J., Cui,P., Han,Y.-J., Xu,H., Cong,L.-J., Huang,X.-G., Wang,X.-L., Jiao,Y.-Z., Wang,B.-J., Wang,Y.-J., Zhang,J.-S.				
Qy 948 GTGTITTAGCTGCTCTGGCTACCAAAAGAAAAAAAGAAAAAAAGAAAAAA 992	TITLE Soybean Expressed Sequence Tags Sequencing				
Db 601 NSSNSNANNNNAAAAVNAANRAAANRAAASAAAAAAAGAAAAAA 557	JOURNAL Unpublished				
RESULT 48	COMMENT Contact: Chen S.Y.				
LOCUS CB615918/c	LOCATION Plant Biotechnology Laboratory				
DEFINITION AGNNNUC:URRG1-00010-E11-A urrg1 (14046) Rattus norvegicus cDNA.	INSTITUTE Institute of Genetics and Developmental Biology, CAS, China				
ACCESSION CE615918	ADDRESS Datun road, Beijing 100101, China				
VERSION CE615918.1	TELEPHONE 86-10-64886859				
KEYWORDS Rattus norvegicus (Norway rat)	FAX 86-10-64873428				
ORGANISM Mammalia; Eutheria; Rodentia; Chordata; Craniata; Vertebrata; Euteleostomi; Murinae; Murinae;	EMAIL sychen@genetics.ac.cn				
SOURCE Rattus norvegicus	SEQ_PRIMER T7 primer.				
VERSION CE615918.1	LOCATION/QUALIFIERS 1.-626				
KEYWORDS Rattus norvegicus	FEATURES SOURCE /organism="Glycine max"				
ORGANISM Mammalia; Eutheria; Rodentia; Chordata; Craniata; Vertebrata; Euteleostomi; Murinae; Murinae;	/mol_type="mRNA"				
SOURCE Rattus norvegicus	/clone_lib="Soybean induced by Salicylic Acid"				
VERSION CE615918.1	/note="Vector: pBlueScript SK+; Site 1: EcoR I; Site 2: Xba I; The cDNA library was constructed by He, C-Y from mRNA isolated from two-week seedlings (cultivar Kefeng 1)"				
KEYWORDS Rattus norvegicus	/db_xref="Taxon:3847"				
ORGANISM Mammalia; Eutheria; Rodentia; Chordata; Craniata; Vertebrata; Euteleostomi; Murinae; Murinae;	/tissue_type="Seedling"				
SOURCE Rattus norvegicus	/dev_stage="Two-week seedlings"				
VERSION CE615918.1	/lab_host="X11-Blue MRP" strain"				
KEYWORDS Rattus norvegicus	/clone_lib="Soybean induced by Salicylic Acid"				
ORGANISM Rattus norvegicus	/note="Vector: pBlueScript SK+; Site 1: EcoR I; Site 2: Xba I; The cDNA library was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XbaI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XbaI digestion. The cDNA fragments were directionally cloned into the				
SOURCE Angen, Inc	EcoRI-XbaI restriction site of the pBlueScript vector. The ligated cDNA fragments were transformed into X11-Blue MRP, host cells (Stratagene)."				
VERSION CE615918.1	RESULTS source				
KEYWORDS One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA	source /organism="Rattus norvegicus"				
ORGANISM Rattus norvegicus	/mol_type="mRNA"				
SOURCE Angen EST Program	/db_xref="taxon:10116"				
VERSION CE615918.1	/clone_lib="urrg1 (14046)"				
KEYWORDS Angen Rat EST Program	/note="Vector: pTT3D-PAC"				
ORGANISM Rattus norvegicus	RESULTS source				
SOURCE Angen, Inc	source /organism="Rattus norvegicus"				
VERSION CE615918.1	/mol_type="mRNA"				
KEYWORDS One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA	/db_xref="taxon:10116"				
ORGANISM Rattus norvegicus	/clone_lib="urrg1 (14046)"				
SOURCE Angen, Inc	/note="Vector: pTT3D-PAC"				

Result No.	Score	Query	Match Length	DB_ID	Description
1	992	100.0	992	4	US-09-995-243-222
C	2	4.7	87350	3	US-08-781-891-79
C	3	4.7	87350	4	US-09-618-166-79
C	4	4.7	87543	4	US-09-791-211-3
C	5	43.6	4.4	289	3 US-09-007-005-17
C	6	43.6	4.4	289	3 US-09-244-96-17
C	7	41.4	4.2	7218	1 US-0-232-63-14
C	8	40.6	4.1	480	4 US-09-252-991A-11640
C	9	40.6	4.1	1572	4 US-09-232-991A-11799
C	10	40.6	4.1	2022	4 US-09-232-991A-11716
C	11	40.2	4.1	1332	4 US-09-252-991A-10993
C	12	40.2	4.1	1656	4 US-09-252-991A-11200
C	13	40.2	4.1	2217	4 US-09-252-991A-11190
C	14	37.8	4.8	4403765	3 US-09-103-840A-2
C	15	37.8	3.8	4411529	3 US-09-103-840A-1
C	16	37.6	3.8	1046	4 US-09-520-714-21
C	17	37.6	3.8	1396	1 US-08-123-161A-11
C	18	37.6	3.8	1395	1 US-08-433-278-11
C	19	37.6	3.8	1992	4 US-09-252-991A-5567
C	20	37.4	3.8	4403765	3 US-08-997-080-179
C	21	37.4	3.8	520	2 US-08-997-362-179
C	22	37.4	3.8	520	3 US-09-055-852-179
C	23	37.4	3.8	520	4 US-09-324-542-179
C	24	37.4	3.8	520	4 US-09-055-426-179
C	25	37.4	3.8	570	3 US-09-055-855-202
C	26	37.4	3.8	570	4 US-09-205-426-202
C	27	36.6	3.7	1273	3 US-09-319-892-3
C	28	3.7	3.6	1273	4 US-09-727-238-3
C	29	3.6	3.6	29	4 US-09-252-991A-11102
C	30	3.6	3.7	1116	4 US-09-252-991A-11158
C	31	3.6	3.7	1584	4 US-09-252-991A-11094
C	32	3.6	3.7	2028	3 US-09-211-930-12
C	33	3.6	3.7	2028	3 US-09-340-993-12
C	34	3.6	3.7	2028	4 US-09-468-442-12
C	35	3.6	3.6	570	4 US-09-252-991A-5589
C	36	3.6	3.6	1545	4 US-09-252-991A-5553
C	37	3.6	3.6	1914	4 US-09-252-991A-5670
C	38	3.6	3.6	2653	3 US-09-245-041-18
C	39	3.6	3.6	4403765	3 US-09-103-840A-2
C	40	3.6	3.6	4411529	3 US-09-103-840A-1
C	41	35.8	3.6	28958	1 US-08-258-261B-6
C	42	35.8	3.6	28958	1 US-08-456-837-6
C	43	35.8	3.6	28958	1 US-08-457-342-6
C	44	35.8	3.6	28958	1 US-08-458-076A-6
C	45	35.8	3.6	28958	1 US-08-764-232A-4
C	46	35.8	3.6	28958	1 US-08-775-932-11
C	47	35.8	3.6	28958	1 US-08-457-335A-6
C	48	35.8	3.6	28958	1 US-08-759-214-6
C	49	35.8	3.6	28958	3 US-09-058-934-6
C	50	35.8	3.6	49377	1 US-08-764-233A-1
C	51	35.6	3.6	366	4 US-19-75-932-13
C	52	35.6	3.6	387	4 US-09-755-932-11
C	53	35.6	3.6	581	2 US-08-461-03C-1
C	54	35.6	3.6	588	3 US-08-744-138-1
C	55	35.6	3.6	588	3 US-09-241-376-1
C	57	35.4	3.6	294	4 US-09-252-991A-10962
C	58	35.4	3.6	447	4 US-19-022-991A-8928
C	59	35.4	3.6	676	4 US-09-183-861-3
C	60	35.4	3.6	676	4 US-19-022-765-3
C	61	35.4	3.6	676	4 US-19-551-974-3
C	62	35.4	3.6	696	4 US-09-252-991A-9216
C	63	35.4	3.6	1011	4 US-09-222-991A-8814
C	64	35.4	3.6	1125	4 US-09-222-991A-14001
C	65	35.4	3.6	1617	4 US-09-252-991A-14135
C	66	35.4	3.6	1728	4 US-09-252-991A-9152
C	67	35.2	3.5	23573	4 US-09-773-81-6
C	68	35	3.5	588	4 US-09-005-258-64
C	69	35	3.5	1683	3 US-09-247-803-11
C	70	34.8	3.5	1646	4 US-09-443-01A-19
C	71	34.8	3.5	1969	1 US-07-337-619-28
C	72	34.8	3.5	1969	3 US-08-029-170-28
C	73	34.6	3.5	816	4 US-09-552-991A-11434
C	74	34.6	3.5	1116	4 US-09-522-714-31
C	75	34.6	3.5	1230	4 US-09-461-325-80
C	76	34.6	3.5	1254	4 US-09-252-991A-11478
C	77	34.6	3.5	1524	4 US-09-252-991A-11455
C	78	34.6	3.5	2028	4 US-09-252-991A-11416
C	79	34.6	3.5	2346	4 US-09-015-839-5
C	80	34.2	3.4	3471	5 PCT-US93-00227-2
C	81	34.6	3.5	3471	5 PCT-US93-00227-2
C	82	34.6	3.5	2350	3 US-08-704-71A-10
C	83	34.6	3.5	3530	4 US-09-521-220-10
C	84	34.6	3.5	4190	3 US-08-938-291A-2
C	85	34.6	3.5	4190	4 US-09-193-503B-5
C	86	34.2	3.4	423	4 US-08-470-179-186
C	87	34.2	3.4	1863	4 US-09-252-991A-32225
C	88	34.2	3.4	2154	4 US-09-052-991A-3346
C	89	34.2	3.4	9208	4 US-09-068-061-1
C	90	34	3.4	2461	1 US-08-282-141-1
C	91	34	3.4	3402	1 US-09-252-991A-1374
C	92	34	3.4	3687	4 US-09-252-991A-1193
C	93	34	3.4	4266	4 US-09-252-991A-1234
C	94	33.8	3.4	1518	3 US-09-240-151-2
C	95	33.8	3.4	1518	3 US-09-591-435-2
C	96	33.8	3.4	1536	4 US-09-252-991A-11652
C	97	33.8	3.4	1656	4 US-09-252-991A-866-63
C	98	33.8	3.4	1761	4 US-09-252-991A-11704
C	99	33.8	3.4	1767	4 US-09-252-991A-1614
C	100	33.8	3.4	2118	4 US-09-252-991A-11782

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

** Sequence 1, App1

C 101	33.8	US-08-313-185-57	Sequence 57, Appli	APPLICANT: Kljavin,Ivar J.
C 102	33.8	US-09-082-614A-57	Sequence 57, Appli	APPLICANT: Napier,Mary A.
C 103	33.6	US-08-804-432A-9	Sequence 9, Appli	APPLICANT: Pan,James
C 104	33.6	US-08-720-229-9	Sequence 9, Appli	APPLICANT: Paoni,Nicholas F.
C 105	33.6	US-09-185-501B-12	Sequence 12, Appli	APPLICANT: Roy, Margaret Ann
C 106	33.6	US-09-122-936-60	Sequence 60, Appli	APPLICANT: Stewart,Timothy A.
C 107	33.4	US-08-948-113D-11	Sequence 11, Appli	APPLICANT: Tumas, Daniel
C 108	33.4	US-09-252-991A-14223	Sequence 14223, A	APPLICANT: Watanabe, Colin K.
C 109	33.4	US-08-385-117	Sequence 117, Appli	APPLICANT: Williams, P. Mickey
C 110	33.4	US-08-267-803B-8	Sequence 8, Appli	APPLICANT: Wood, William L.
C 111	33.4	US-09-041-986-16	Sequence 16, Appli	APPLICANT: Zhang, Zemin
C 112	33.2	US-09-52-991A-12679	Sequence 12679, A	TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
C 113	33.2	US-09-232-991A-13279	Sequence 13279, A	Acids Encoding the Same
C 114	33.2	US-09-252-991A-8389	Sequence 8589, Ap	FILE REFERENCE: P2730P1C13
C 115	33.2	US-09-52-991A-12822	Sequence 12822, A	CURRENT APPLICATION NUMBER: US-09/996,243
C 116	33.2	US-08-184-632-1	Sequence 1, Appli	CURRENT FILING DATE: 2001-11-14
C 117	33.2	US-07-951-715A-3	Sequence 3, Appli	PRIOR APPLICATION NUMBER: 60/049787
C 118	33.2	US-08-45-48A-3	Sequence 3, Appli	PRIOR FILING DATE: 1997-06-16
C 119	33.2	US-08-459-59A-3	Sequence 3, Appli	PRIOR APPLICATION NUMBER: 60/062250
C 120	33.2	US-08-459-59A-3	Sequence 3, Appli	PRIOR FILING DATE: 1997-10-17
C 121	33.2	US-08-459-59A-3	Sequence 3, Appli	PRIOR APPLICATION NUMBER: 60/065186
C 122	33.2	US-09-547-422-3	Sequence 3, Appli	PRIOR FILING DATE: 1997-11-12
C 123	33.2	US-08-492-027A-7	Sequence 7, Appli	PRIOR APPLICATION NUMBER: 60/065311
C 124	33.2	US-08-948-589A-9	Sequence 9, Appli	PRIOR FILING DATE: 1997-11-13
C 125	33.2	US-09-188-46-9	Sequence 9, Appli	PRIOR APPLICATION NUMBER: 60/066770
C 126	33.2	US-09-397-22-8A-9	Sequence 9, Appli	PRIOR FILING DATE: 1997-11-24
C 127	33.2	US-09-252-991A-8656	Sequence 8656, Ap	PRIOR APPLICATION NUMBER: 60/075945
C 128	33.2	US-09-027-137-2	Sequence 2, Appli	PRIOR FILING DATE: 1998-02-25
C 129	33.2	US-08-344-41-2	Sequence 2, Appli	PRIOR APPLICATION NUMBER: 60/078910
C 130	33.2	US-07-551-715A-2	Sequence 2, Appli	PRIOR FILING DATE: 1998-03-20
C 131	33.2	US-08-951-715A-4	Sequence 2, Appli	PRIOR APPLICATION NUMBER: 60/083322
C 132	33.2	US-07-551-715A-8	Sequence 8, Appli	PRIOR FILING DATE: 1998-04-28
C 133	33.2	US-08-459-448A-2	Sequence 2, Appli	PRIOR APPLICATION NUMBER: 60/084600
C 134	33.2	US-08-459-448A-4	Sequence 4, Appli	PRIOR FILING DATE: 1998-05-07
C 135	33.2	US-08-459-504B-8	Sequence 8, Appli	PRIOR APPLICATION NUMBER: 60/087106
C 136	33.2	US-08-459-555A-2	Sequence 4, Appli	PRIOR FILING DATE: 1998-05-28
C 137	33.2	US-08-459-555A-4	Sequence 4, Appli	PRIOR APPLICATION NUMBER: 60/087607
C 138	33.2	US-08-459-595A-8	Sequence 8, Appli	PRIOR FILING DATE: 1998-06-02
C 139	33.2	US-08-459-504B-2	Sequence 2, Appli	PRIOR APPLICATION NUMBER: 60/087609
C 140	33.2	US-08-459-504B-4	Sequence 4, Appli	PRIOR FILING DATE: 1998-06-02
C 141	33.2	US-08-459-504B-8	Sequence 8, Appli	PRIOR APPLICATION NUMBER: 60/087759
C 142	33.2	US-08-459-444-2	Sequence 4, Appli	PRIOR FILING DATE: 1998-06-02
C 143	33.2	US-08-459-444-4	Sequence 4, Appli	PRIOR APPLICATION NUMBER: 60/087827
C 144	33.2	US-08-459-444-8	Sequence 8, Appli	PRIOR FILING DATE: 1998-06-03
C 145	33.2	US-08-459-444-8	Sequence 8, Appli	PRIOR APPLICATION NUMBER: 60/088021
C 146	33.2	US-09-053-549-3	Sequence 5, Appli	PRIOR FILING DATE: 1998-06-04
C 147	33.2	US-09-547-122-3	Sequence 2, Appli	PRIOR APPLICATION NUMBER: 60/088029
C 148	33.2	US-09-547-122-4	Sequence 4, Appli	PRIOR FILING DATE: 1998-06-04
C 149	33.2	US-09-547-122-8	Sequence 8, Appli	PRIOR APPLICATION NUMBER: 60/088025
C 150	33.2	US-07-951-715A-10	Sequence 10, Appli	PRIOR FILING DATE: 1998-06-04

ALIGNMENTS

RESULT¹
US-09-996-243-222
/ General Information:
/ Patent No. 6478825

/ APPLICANT: Ashkenazi,Avi J.
/ APPLICANT: Baker,Kevin P.
/ APPLICANT: Botstein,David
/ APPLICANT: Desnoyers,Luc
/ APPLICANT: Eaton,Dan L.
/ APPLICANT: Ferrara,Napoleone
/ APPLICANT: Fong,Sherman
/ APPLICANT: Gerber,Hanspeter
/ APPLICANT: Gerritsen,Mary E.
/ APPLICANT: Goddard,Audrey
/ APPLICANT: Godowski,Paul J.
/ APPLICANT: Grimaldi,J.Christopher
/ APPLICANT: Gurney,Austin L.

Db 421 AGCTGTGGTCAAGCCACTGTGTCAGCTGGAGACAGAGGG 480
 Qy CAGGCCCAAGGTTGGAGATGATCTGGCAGGCCTCTGGGAGCC 540
 Db 481 CAGGCCCAAGGTTGGAGATGATCTGGCAGGCCTCTGGGAGCC 540
 Qy 541 GCTGTGTCAGGAGGATGGCAGAGACCATGGCAGGCCTG 600
 Db 541 GCTGTGTCAGGAGGATGGCAGAGACCATGGCAGGCCTG 600
 Qy 601 CCAACTCTCTTCCTCGGAGCATGGACTGGTTCTGGGTCAGGG 660
 Db 601 CCAACTCTCTTCCTCGGAGCATGGACTGGTTCTGGGTCAGGG 660
 Qy 661 AGCCAAATGTCAGGACAGGCCCTCACTGAGTGGCAGAGATGG 720
 Db 661 AGCCAAATGTCAGGACAGGCCCTCACTGAGTGGCAGAGATGG 720
 Qy 721 AGGACTGGAGGTGGTCCCTGGAGGAGCCATCTTGCTTGGCCT 780
 Db 721 AGGACTGGAGGTGGTCCCTGGAGGAGCCATCTTGCTTGGCCT 780
 Qy 781 GCGGTCTGAGTGAAGGAGTTGGGGTTCAAGGATAAGGGATGGG 840
 Db 781 GCGGTCTGAGTGAAGGAGTTGGGGTTCAAGGATAAGGGATGGG 840
 Qy 841 GCAAGGAGGAGCCATGTAGATGAGCCGTCAAGAGCCAGGACTGCA 900
 Db 841 GCAAGGAGGAGCCATGTAGATGAGCCGTCAAGAGCCAGGACTGCA 900
 Qy 901 GCCCATAGGTGCACTGTGTTATTGGAGTTCTGAAATGAGTGTGTT 960
 Db 901 GCCCATAGGTGCACTGTGTTATTGGAGTTCTGAAATGAGTGTGTT 960
 Qy 961 TCTTGGCCACAAAAA 992
 Db 961 TCTTGGCCACAAAAA 992

RESULT 2
 US-08-781-891-79/c
 ; Sequence 79, Application US/08781891
 ; Patent No. 6090620
 ; GENERAL INFORMATION:
 ; APPLICANT: Fu, Ying-Hui
 ; APPLICANT: Yu, Chang-En
 ; APPLICANT: Oshima, Junko
 ; APPLICANT: Mulligan, John T.
 ; APPLICANT: Scheellenberg, Gerald D.
 ; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 ; TITLE OF INVENTION: WERNER'S SYNDROME
 ; NUMBER OF SEQUENCES: 209
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/618,166
 ; FILING DATE: 17-Jul-2000
 ; CLASSIFICATION: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McMasters, David D.
 ; REGISTRATION NUMBER: 33,963
 ; REFERENCE/DOCKET NUMBER: 240052-419C1
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 79:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 97350 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
 ; US-09-618-166-79

Query Match 4.7%; Score 47; DB 3; Length 87350;
 Best Local Similarity 78.9%; Pred. No. 0.045;
 Matches 56; Conservative 0; MisMatches 15; Indels 0; Gaps 0;
 SEQUENCE CHARACTERISTICS:
 LENGTH: 87350 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-781-891-79

Db 421 AGCTGTGGTCAAGCCACTGTGTCAGCTGGAGACAGAGGG 480
 Qy CAGGCCCAAGGTTGGAGATGATCTGGCAGGCCTCTGGGAGCC 540
 Db 481 CAGGCCCAAGGTTGGAGATGATCTGGCAGGCCTCTGGGAGCC 540
 Qy 541 GCTGTGTCAGGAGGATGGCAGAGACCATGGCAGGCCTG 600
 Db 541 GCTGTGTCAGGAGGATGGCAGAGACCATGGCAGGCCTG 600
 Qy 601 CCAACTCTCTTCCTCGGAGCATGGACTGGTTCTGGGTCAGGG 660
 Db 601 CCAACTCTCTTCCTCGGAGCATGGACTGGTTCTGGGTCAGGG 660
 Qy 661 AGCCAAATGTCAGGACAGGCCCTCACTGAGTGGCAGAGATGG 720
 Db 661 AGCCAAATGTCAGGACAGGCCCTCACTGAGTGGCAGAGATGG 720
 Qy 721 AGGACTGGAGGTGGTCCCTGGAGGAGCCATCTTGCTTGGCCT 780
 Db 721 AGGACTGGAGGTGGTCCCTGGAGGAGCCATCTTGCTTGGCCT 780
 Qy 781 GCGGTCTGAGTGAAGGAGTTGGGGTTCAAGGATAAGGGATGGG 840
 Db 781 GCGGTCTGAGTGAAGGAGTTGGGGTTCAAGGATAAGGGATGGG 840
 Qy 841 GCAAGGAGGAGCCATGTAGATGAGCCGTCAAGAGCCAGGACTGCA 900
 Db 841 GCAAGGAGGAGCCATGTAGATGAGCCGTCAAGAGCCAGGACTGCA 900
 Qy 901 GCCCATAGGTGCACTGTGTTATTGGAGTTCTGAAATGAGTGTGTT 960
 Db 901 GCCCATAGGTGCACTGTGTTATTGGAGTTCTGAAATGAGTGTGTT 960
 Qy 961 TCTTGGCCACAAAAA 992
 Db 961 TCTTGGCCACAAAAA 992

RESULT 3
 US-09-618-166-79/c
 ; Sequence 79, Application US/09618166
 ; Patent No. 6583112
 ; GENERAL INFORMATION:
 ; APPLICANT: Fu, Ying-Hui
 ; APPLICANT: Yu, Chang-En
 ; APPLICANT: Oshima, Junko
 ; APPLICANT: Mulligan, John T.
 ; APPLICANT: Scheellenberg, Gerald D.
 ; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 ; NUMBER OF SEQUENCES: 209
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed Intellectual Property Law Group
 ; STREET: 701 Fifth Avenue, Suite 6300
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/618,166
 ; FILING DATE: 17-Jul-2000
 ; CLASSIFICATION: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McMasters, David D.
 ; REGISTRATION NUMBER: 33,963
 ; REFERENCE/DOCKET NUMBER: 240052-419C1
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 79:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 97350 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
 ; US-09-618-166-79

Query Match 4.7%; Score 47; DB 4; Length 87350;
 Best Local Similarity 78.9%; Pred. No. 0.045;
 Matches 56; Conservative 0; MisMatches 15; Indels 0; Gaps 0;
 SEQUENCE CHARACTERISTICS:
 LENGTH: 97350 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-618-166-79

Db 421 AGCTGTGGTCAAGCCACTGTGTCAGCTGGAGACAGAGGG 480
 Qy CAGGCCCAAGGTTGGAGATGATCTGGCAGGCCTCTGGGAGCC 540
 Db 481 CAGGCCCAAGGTTGGAGATGATCTGGCAGGCCTCTGGGAGCC 540
 Qy 541 GCTGTGTCAGGAGGATGGCAGAGACCATGGCAGGCCTG 600
 Db 541 GCTGTGTCAGGAGGATGGCAGAGACCATGGCAGGCCTG 600
 Qy 601 CCAACTCTCTTCCTCGGAGCATGGACTGGTTCTGGGTCAGGG 660
 Db 601 CCAACTCTCTTCCTCGGAGCATGGACTGGTTCTGGGTCAGGG 660
 Qy 661 AGCCAAATGTCAGGACAGGCCCTCACTGAGTGGCAGAGATGG 720
 Db 661 AGCCAAATGTCAGGACAGGCCCTCACTGAGTGGCAGAGATGG 720
 Qy 721 AGGACTGGAGGTGGTCCCTGGAGGAGCCATCTTGCTTGGCCT 780
 Db 721 AGGACTGGAGGTGGTCCCTGGAGGAGCCATCTTGCTTGGCCT 780
 Qy 781 GCGGTCTGAGTGAAGGAGTTGGGGTTCAAGGATAAGGGATGGG 840
 Db 781 GCGGTCTGAGTGAAGGAGTTGGGGTTCAAGGATAAGGGATGGG 840
 Qy 841 GCAAGGAGGAGCCATGTAGATGAGCCGTCAAGAGCCAGGACTGCA 900
 Db 841 GCAAGGAGGAGCCATGTAGATGAGCCGTCAAGAGCCAGGACTGCA 900
 Qy 901 GCCCATAGGTGCACTGTGTTATTGGAGTTCTGAAATGAGTGTGTT 960
 Db 901 GCCCATAGGTGCACTGTGTTATTGGAGTTCTGAAATGAGTGTGTT 960
 Qy 961 TCTTGGCCACAAAAA 992
 Db 961 TCTTGGCCACAAAAA 992

RESULT 2
 US-08-781-891-79/c
 ; Sequence 79, Application US/08781891
 ; Patent No. 6090620
 ; GENERAL INFORMATION:
 ; APPLICANT: Fu, Ying-Hui
 ; APPLICANT: Yu, Chang-En
 ; APPLICANT: Oshima, Junko
 ; APPLICANT: Mulligan, John T.
 ; APPLICANT: Scheellenberg, Gerald D.
 ; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 ; TITLE OF INVENTION: WERNER'S SYNDROME
 ; NUMBER OF SEQUENCES: 209
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/781,891
 ; FILING DATE: 27-Dec-1996
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: No. 6090620enburg Ph.D., Carol
 ; REGISTRATION NUMBER: 39,317
 ; REFERENCE/DOCKET NUMBER: 240052-419
 ; TELECOMMUNICATION INFORMATION:
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
 ; US-09-618-166-79

Query Match 4.7%; Score 47; DB 4; Length 87350;
 Best Local Similarity 78.9%; Pred. No. 0.045;
 Matches 56; Conservative 0; MisMatches 15; Indels 0; Gaps 0;
 SEQUENCE CHARACTERISTICS:
 LENGTH: 97350 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-618-166-79

RESUIT⁴
 US-09-791-211-3/C
 / Sequence 3, Application US/09791211
 / Patent No. 6448080
 / GENERAL INFORMATION:
 / APPLICANT: Donna T. Ward
 / TITLE OF INVENTION: ANTISENSE MODULATION OF WEN EXPRESSION
 / FILE REFERENCE: RTS-0205
 / CURRENT APPLICATION NUMBER: US/09/791,211
 / FILING DATE: 2001-02-23
 / NUMBER OF SEQ ID NOS: 90
 / SEQ ID NO 3
 / LENGTH: 87543
 / TYPE: DNA
 / FEATURE:
 / NAME/KEY: unsure
 / LOCATION: 7421
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 7427
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 11609
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 12605
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 12742
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 29370
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 12942
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 29979
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 30136
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 30140
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 31592
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 31206
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 33095
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 33160
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 34066
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 34072
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 34916
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 39020
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 42164
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 42459
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 46808
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 46823
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 46826
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 47291
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 52786
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 52787
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 53384
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 54684
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 59215
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 59235
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 66614
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 68697
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 68718
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure
 / LOCATION: 68733
 / OTHER INFORMATION: unknown
 / NAME/KEY: unsure

CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
COUNTY: VA
ZIP: 22313-0299

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEFAX: (703) 836-9300
TELELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZGPT-F1S

Best Local Similarity 44.4%; Pred. No. 0.44; Mismatches 0; Indels 204; Gaps 0;

Matches 163; Conservative 0; Gaps 0;

Qy 260 GAACATAGTGGCCAGAACGAGGTGGTGAAGACCAAGAGCCACTTCACCTCAA 319
 Db 1173 GGAGATAAGAAGAACCTATGCCAGGGCTCCCTCAACCTCAA 1232
 Qy 320 CGTCACACTAACAGTCCAGACCTGCTCACCTACTTGCGGCTCGG 1237
 Db 1233 CGTTCGGACCTGACATCCAACTTGACAGAACCTATTCGGCGCTC 1292
 Qy 380 AGGTGCCATTGTGGACAGTGGCAGGGTACAGATGCACTGGAGCTGTGGTCAACGCCATG 439
 Db 1293 CGGGGGATGGCCACGGGATGAAATCTCAAGATGCCATCGGCATCGG 1352
 Qy 440 GTCTGAGCTGGGGCAAACTTCACTTGAGGAGACAGGGCAGGGGGTGGATG 499
 Db 1353 CAGCGAGTTGCGGCGCAGCGAGTGTAGAGTCAACTAGCGGGGCGCCAGGAGA 1412
 Qy 500 GATCCCATGAGGCTGGCCAGGCCAACCTTACACAGGGAAAGGATGG 559
 Db 1413 GATCCCATGAGGCTGGCCAGGCCAACCTTACACAGGGAAAGGATGG 1472
 Qy 560 GCAGGTTCCACCTGTGACCAAGGACCATGCCAACCTTCCCTGTCGCC 619
 Db 1473 CATGGTGAGATGTGATGGGACTAACAGAGAGACGGGGTGGACGGTGGCACTGGA 1532
 Qy 620 GAGCCAG 626
 Db 1533 CAACCCG 1539

RESULT 11
 US-09-252-991A-10993
 ; Sequence 10993, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEAR ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10993
 ; LENGTH: 1332
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-10993

Query Match Score 40.2; DB 4; Length 1332;
 Best Local Similarity 49.8%; Pred. No. 0.53; Indels 0; Gaps 0;
 Matches 102; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 218 CCAGCCACCAACGCCCATCACCTATTCCTCTGTGGAAACCAAAACTCAAGGTGGCAA 277
 Db 826 CCCGCCAACCGCAAGGATGCGGGATTCGGGATTCGGGAATACCAACTCGTGAACCGA 885
 Qy 278 GAAAGTGTGTAAGACCCACGAGCGGGCTCCPTCAACTCAAGTCAAGTCAC 337
 Db 886 GCGGATGCGATGTGCGGTGACGCCACATCGGCTGAGGACCATGACCC 945
 Qy 338 TCCAGACCTGCTCACCTACTCTGCCCCGGCTCTCACCTAGGTGCCATGTGGCAG 397
 Db 946 GCGGATCAGGGGAGGGGGATCTGGGATTCAGGGTCAACCTAGGTGCCATGTGGCAG 1005
 Qy 398 TGCGAGGTACAGATGGACTGGAG 422
 Db 1006 CGCGAGGGGGATCTGGCGGAG 1030

RESULT 12
 US-09-252-991A-11258/c
 ; Sequence 11258, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEAR ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142

Query Match Score 40.6; DB 4; Length 2022;
 Best Local Similarity 44.4%; Pred. No. 0.48; Mismatches 204; Indels 0; Gaps 0;

Matches 163; Conservative 0; Gaps 0;

Qy 260 GAACATAGTGGCCAGAACGAGGTGGTGAAGACCAAGAGCCACTTCACCTCAA 319
 Db 1078 GGAGATAAGAAGAACCTATGCCAGGGCTCCCTCAACCTCAA 1137
 Qy 320 CGTCACACTAACAGTCCAGACCTGCTCACCTACTTGCGGCTCGG 1197
 Db 1138 CGTTGCGACCTGCGACAATCGCAACTGAGGTTGCTGCCAGGCTGCTACATTCGGCCCTCGG 1257
 Qy 380 AGGTGCCATTGTGGACAGTGGCCAGGCTAACATGCACTGGAGCTGTGTCACAGGCAAGGCGACT 439
 Db 1198 CGGGGGATGGGGCAACCGGAGGATCAACTCTGAGGAGCCATTGGCGGCTCGG 1257
 Qy 440 GTCTGAGCTGGGGCAACTTCACCTGCGGAGAGGGCAAGGGTGGAGAT 499
 Db 1258 CAGCAGITGGGGCGCAACGAGCTAACATAGGAGTCAACTAACGGAGGA 1317

RESULT 14
US-09-103-840A-2/C
i SEQ ID NO 11258
i LENGTH: 1656
i TYPE: DNA
i ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-11258

Query Match	4.1%	Score 40.2;	DB 4;	Length 1656;
Best Local Similarity	49.8%;	Pred. No.	0.57;	
Matches 102;	Conservative 0;	Mismatches	103;	Indels 0;
Gaps 0;				

Qy 218 CCAGGCCACCA CGCCCATACCTATTCCCTTGAAAGAACATCAAGTGGCCAA 277
 Db 570 CCGCCACCGCAGGATCGGGATCCGGGATTCACCTCGGTGACCGA 511

Qy 278 GAAAGTGGTGAAGACCCAGGAGCCAGGCTCTCTAACCTCAGTCAACTCAAGTCAGTCAG 337
 Db 510 GCCATGGGATGGTGCCTGACCGCCACCATAGGGTGTAGCGGACCATGACCC 451

Qy 338 TCCAGACCTGCTCACCTACTCMCCGGGGTCTCCACCTCAGTGGCCAGTGGCAAG 397
 Db 450 GCGGATCACGGCCAGGGCTGGTACCGCTGCTCCAGAACCTGGGTACCGTGAACAC 391

Qy 398 TGCAGGTTACAGATGACAGTGGAG 422
 Db 390 CGCGAGGGGGGATCAGCCGGAG 366

Qy 398 TGCAGGTTACAGATGACAGTGGAG 422
 Db 390 CGCGAGGGGGGATCAGCCGGAG 366

RESULT 13
US-09-252-991A-11190/C
i Sequence 11190, Application US/09252991A
i Patent No. 6551795
i GENERAL INFORMATION:
i APPLICANT: Marc J. Rubenstein et al.
i TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
i TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
i FILE REFERENCE: 1-07196-136
i CURRENT APPLICATION NUMBER: US/09/252, 991A
i CURRENT FILING DATE: 1999-02-18
i PRIOR APPLICATION NUMBER: US 60/074, 788
i PRIOR FILING DATE: 1998-02-18
i PRIOR APPLICATION NUMBER: US 60/094, 190
i PRIOR FILING DATE: 1998-07-27
i NUMBER OF SEQ ID NOS: 33142
i SEQ ID NO 11190
i LENGTH: 2217
i TYPE: DNA
i ORGANISM: Pseudomonas aeruginosa
i FEATURE:
i NAME/KEY: unsure
i LOCATION: (12217)
i OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
 US-09-252-991A-11190

Query Match

4.1%	Score 40.2;	DB 4;	Length 2217;
Best Local Similarity	49.8%;	Pred. No.	0.63;
Matches 102;	Conservative 0;	Mismatches	103;
Indels 0;			
Gaps 0;			

Qy 218 CGAGGCCACCA CGCCCATACCTATTCCCTTGAAAGAACATCAAGTGGCCAA 277
 Db 778 CCCGCCACCGGAGGATCGGGATTCGGGATTCACCTCGGTGACCGA 719

Qy 278 GAGGGTGTGAAGACCCAGGAGCCCTGCTCAACCTCAAGTCAACTCAAGTCAG 337
 Db 718 CGCGATGGGTGCTGGGATCCGGGATTCAGGTCTAGGGACCATGACCC 659

Qy 338 TCCAGACCTGCTCACCTACTCTGCGGGCTCCACCTCAAGTGGCCATGTGACG 397
 Db 658 GCGGATAGGGGGAGGGCTGCTAGGCTGCTGGGATTCAGGTGAACAC 599

Qy 398 TGCCAGGCTATAAGATGCACTGGAG 422
 Db 598 CGCGAGGGGGCGGATCTGCCCGAG 574

RESULT 15
US-09-103-840A-1/C
i Sequence 1, Application US/09103840A
i Patent No. 6294328
i GENERAL INFORMATION:
i APPLICANT: FLEISCHMAN, Robert D.
i APPLICANT: WHITE, Owen R.
i APPLICANT: FRASER, Claire M.
i APPLICANT: VENTER, John C.
i TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
i CURRENT APPLICATION NUMBER: 24366-20007.00
i FILE REFERENCE: 24366-20007.00
i CURRENT FILING DATE: 1998-06-24
i NUMBER OF SEQ ID NOS: 2
i SOFTWARE: PatentIn Ver. 2.1
i SEQ ID NO 1
i LENGTH: 4403765
i TYPE: DNA
i ORGANISM: Mycobacterium tuberculosis
i FEATURE:
i OTHER INFORMATION: CDC 1551
i OTHER INFORMATION: "n" bases at various positions throughout the sequence
 US-09-103-840A-2

Query Match

3.8%	Score 37.8;	DB 3;	Length 4403765;
Best Local Similarity	47.3%;	Pred. No.	0.34;
Matches 114;	Conservative 0;	Mismatches	127;
Indels 0;			
Gaps 0;			

Qy 338 TCCAGACCTGCTCACCTACTCTGCGGGCTCCACCTCGGTGCGACAG 397
 Db 284113 TTCCGACATTCCTGGAATCCCTCGGGCTACCGGGCCCGAGGCGATCCGCAC 284054

Qy 398 TGCCAGGCTAGATGCACTGGAGTGTGTCAGTCAGCTGCGGCAA 457
 Db 284053 CGCCCTGACCCGCCCGCCAGGGTGTGTCAGTCAGCTGCGGCAA 283994

Qy 458 CTTCACTCTGAGGAAAGAGGGCGAGGCCAGGTTGCAAGGCTGCTGCTGCGGCT 517
 Db 283993 CCTGGCGGGCCACCGTGTGACCGGCTGCGGCTGCGGCTGCGGCT 283934

Qy 518 GGGCACCCCACCTATCAACACGGCTGATGGTAGGTCCACCTGCGACA 577
 Db 283933 GGCCCCATCCGACGGTGTGGCCATCAACCTGGGAGGCGCCAGGGCTGCGACA 283874

Qy 578 G 578
 Db 283873 G 283873

Db 1309 GGGTGGGTAGAGTGTGAGTAAGGACATTAGAATAATATCTGGCTCTGCCTCA 1368 ; Patent No. 6551795
 Qy 969 CAATAAAATTTAAAAAAATTTAAAAAA 992 ; GENERAL INFORMATION:
 Db 1369 CCATAAAATTTAAAAAAATTTAAAAAA 1392 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PATENT NO.: 586073 ; PRIOR FILING DATE: 1998-02-18
 ; GENERAL INFORMATION: ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; APPLICANT: Campbell, Kevin P. ; PRIOR FILING DATE: 1998-07-27
 ; APPLICANT: Ibraghimov, Oxana B. ; NUMBER OF SEQ ID NOS: 33142
 ; APPLICANT: Ervasti, James M. ; SEQ ID NO: 5567
 ; TITLE OF INVENTION: NUCLEAR ACID ENCODING DYSTROPHIN-ASSOCIATED ; LENGTH: 1992
 ; TITLE OF INVENTION: PROTEIN ; TYPE: DNA
 ; NUMBER OF SEQUENCES: 15 ; ORGANISM: Pseudomonas aeruginosa
 ; CORRESPONDENCE ADDRESS: US-09-252-991A-5567

Query Match Score 3.8%; Best Local Similarity 50.6%; DB 4; Length 1992;
 Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 515 CTCGGCAGGCCACCTATCACCAACCCGTATCGGAAAGATGGCCAGGTCACTTGCA 574
 Db 1569 CTCTGTCGCCGGCGCCGGCGCTGATGACAGATGAGCTGGCTGGTTAGAGGGTTAGAA 1510
 Qy 575 GCAGAGACCATGCCAGGGAGCTCCACTCTTCCTCTGCCGAGCCAGACATGGGA 634
 Db 1509 GAGCAGTCGGGACGGGGTCAGGGCCATCTGTGGAGCCGGCTTCGTTGATG 1450
 Qy 635 CTGGTTCTGGGCCAGGCTGCAACAGCCATGTCAGAGCCGCTTAAGGGT 694
 Db 1449 CTGCTGCTGGAGCAGGCTTTCGAGGAATATCCAGGGAGGGACGGTGTG 1390

RESULT 18
 US-08-483-278-11
 Sequence 11, Application US/08483278
 Patent No. 586073
 GENERAL INFORMATION:
 APPLICANT: Campbell, Kevin P.
 APPLICANT: Ibraghimov, Oxana B.
 APPLICANT: Leveille, Cynthia J.
 TITLE OF INVENTION: NUCLEAR ACID ENCODING DYSTROPHIN-ASSOCIATED
 TITLE OF INVENTION: PROTEIN
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Kevin M. Farrell, P.C.
 STREET: P.O. Box 999
 CITY: York Harbor
 STATE: ME
 COUNTRY: USA
 ZIP: 03911
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/483,278
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/123,161
 FILING DATE: 16-SEP-93
 ATTORNEY/AGENT INFORMATION:
 NAME: Farrell, Kevin M.
 REGISTRATION NUMBER: 35,505
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (207) 363-0558
 TELEFAX: (207) 363-0528
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1396 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 4..1164

Query Match Score 3.8%; Best Local Similarity 65.5%; DB 1; Length 1396;
 Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

US-08-483-278-11

Query Match Score 3.8%; Best Local Similarity 50.6%; DB 4; Length 1992;
 Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 909 GCGTGCAGTCGTGTTGGAGTTATGCAAATGAGTGTGTTAGCTGCTCTGCTCA 968
 Db 1309 GCGTGGGTGAGAGTGTGGAGTAAAGGACATTAGAATAATCTGGCTCTGCTCA 1368

Qy 969 CAATAAAATTTAAAAAAATTTAAAAAA 992
 Db 1369 CCATAAAATTTAAAAAAATTTAAAAAA 1392

RESULT 19
 US-09-252-991A-5567/c
 Sequence 5567, Application US/09252991A
 ; INFORMATION FOR SEQ ID NO: 179:

SEQUENCE CHARACTERISTICS:
 LENGTH: 520 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-997-080-179

Query Match 3.8%; Score 37.4%; DB 2; Length 520;
 Best Local Similarity 48.0%; Pred. No. 1.9;
 Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 374 CACCTCAGGCGCCATGGACAGTCAGATGCCGCTACAGATGCACTGGGGCTGTGTGTCCTCAA 433
 Db 324 CTCCCTGGTGAAGCTGGGGAAACCCGATGATGTTGCCTGGATTGAC 265

QY 434 GCGAGTGTCTAAGCTGGGCCAACCTCACTCTCGGGCAAGGGCGCCAGGGT 493
 Db 264 GCGGAGGTGCGAGTGGATGGCATCTCATGGGGAGTGCTCCCTCTAGGG 205

QY 494 GGAGATGATGATGCCAGGGCTCTCGGGCAACCCATACCAACAGCTGATGGAA 553
 Db 204 CTTGATCACACCATTGGGCTCGGGCACTTGATGTCATGGCTGATGGGT 145

QY 554 GGATGGCAGGTCCACCTGCAACAGCTGATGGCAACACTGATGGCAAGCAG 596
 Db 144 GGAGGGGCCCCGTAGTAGTCATGGTTGACATGGCG 102

RESULT 21
 US-08-997-362-179/C
 Sequence 179, Application US/08997362
 Patent No. 5985287

GENERAL INFORMATION:
 APPLICANT: Tan, Paul
 APPLICANT: Hiyama, Jun
 APPLICANT: Visser, Elizabeth
 APPLICANT: Skinner, Margot
 APPLICANT: Scott, Linda
 APPLICANT: Prestidge, Ross
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
 NUMBER OF SEQUENCES: 194
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Law Offices of Ann W. Speckman
 STREET: 2601 Elliott Avenue, Suite 4185
 CITY: Seattle
 STATE: WA
 COUNTY: USA
 ZIP: 98121

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/997,362
 FILING DATE: August 29, 1996
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
 FILING DATE: June 12, 1997
 APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
 FILING DATE: August 29, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Sleath, Janet
 REGISTRATION NUMBER: 37,007
 REFERENCE/DOCKET NUMBER: 11000.1002c2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 206-269-0565
 TELEFAX: 206-269-0563
 TELEX:
 INFORMATION FOR SEQ ID NO: 179:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 520 base pairs

Query Match 3.8%; Score 37.4%; DB 2; Length 520;
 Best Local Similarity 48.0%; Pred. No. 1.9;
 Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 374 CACCTCAGGCGCCATGGACAGTCAGATGCCGCTACAGATGCACTGGGGCTGTGTGTCCTCAA 433
 Db 324 CTCCCTGGTGAAGCTGGGGAAACCCGATGATGTTGCCTGGATTGAC 265

QY 434 GCGAGTGTCTAAGCTGGGCCAACCTCACTCTCGGGCAAGGGCGCCAGGGT 493
 Db 264 GCGGAGGTGCGAGTGGATGGCATCTCATGGGGAGTGCTCCCTCTAGGG 205

QY 494 GGAGATGATGATGCCAGGGCTCTCGGGCAACCCATACCAACAGCTGATGGAA 553
 Db 204 CTTGATCACACCATTGGGCTCGGGCACTTGATGTCATGGCTGATGGGT 145

QY 554 GGATGGCAGGTCCACCTGCAACAGCTGATGGCAACACTGATGGCAAGCAG 596
 Db 144 GGAGGGGCCCCGTAGTAGTCATGGTTGACATGGCG 102

RESULT 22
 US-09-09-855-179/C
 Sequence 179, Application US/09095855
 Patent No. 6160093

GENERAL INFORMATION:
 APPLICANT: Tan, Paul
 APPLICANT: Visser, Elizabeth
 APPLICANT: Skinner, Margot
 APPLICANT: Prestidge, Ross
 TITLE OF INVENTION: Compounds and Methods for Treatment and Diagnosis of Mycobacterial Infections
 NUMBER OF SEQUENCES: 208
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Law Offices of Ann W. Speckman
 STREET: 2601 Elliott Avenue, Suite 4185
 CITY: Seattle
 STATE: WA
 COUNTY: USA
 ZIP: 98121

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/095,855
 FILING DATE:
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 08/705,347
 FILING DATE: 29-AUG-1996
 APPLICATION NUMBER: 08/873,970
 FILING DATE: 12-JUN-1997
 APPLICATION NUMBER: 08/997,362
 FILING DATE: 23-DEC-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Sleath, Jane
 REGISTRATION NUMBER: 37,007
 REFERENCE/DOCKET NUMBER: 11000.1002c3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 206-269-0565
 TELEFAX: 206-269-0563
 TELEX:
 INFORMATION FOR SEQ ID NO: 179:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 520 base pairs

; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 US-09-855-179

Query Match 3.8%; Score 37.4; DB 3; Length 520;
 Best Local Similarity 48.0%; Pred. No. 1.9; Indels 0; Gaps 0;
 Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 374 CACCTCGGCCATGGACAGTGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
 Db 324 CTCCCTCGTGTGGCTGGCATGACACCCGGATGATGTTGGTGGATGAC 265

Qy 434 GCCAGTGTCTGAGCTGGGGCAACTTCACCTGAGAACAGGGCAGGCCAGGGT 493
 Db 264 GGCGAGGTGGAGTTGGGATGATCCTCGATGAGGGCACTGGTCTCGTAGGG 205

Qy 494 GGAGATGATCTGCAGGGTCCGGGAGGCCACCTATACCAACAGCCATGCGGA 553
 Db 204 CTGTATCACCACTATGCGCCCTGGGAGCTGTGATGTCGAGTCGATCGGGT 145

Qy 554 GGATGGCAGGTGCACTTGAGAGAACATGCCACAGGCG 596
 Db 144 GGAGGCCGCGTAGTGTGATGTCGATGGTGAACATGGCG 102

RESULT 23
 US-09-324-542-179/C
 Sequence 179, Application US/09324542
 Patent No. 6328978

GENERAL INFORMATION:
 ; APPLICANT: Watson, James D.
 ; ATTORNEY: Tar, Paul L.J.
 ; APPLICANT: Prestidge, Ross
 ; TITLE OF INVENTION: Methods and Compounds for the Treatment
 ; TITLE OF INVENTION: Immunologically-Mediated Skin Disorders
 ; FILE REFERENCE: 11000_1007cl
 ; CURRENT APPLICATION NUMBER: US/09/324_542
 ; CURRENT FILING DATE: 1999-06-02
 ; EARLIER APPLICATION NUMBER: US 08/997,080
 ; EARLIER FILING DATE: 1997-12-23
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO: 179
 ; LENGTH: 520
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium vaccae
 US-09-324-542-179

Query Match 3.8%; Score 37.4; DB 4; Length 520;
 Best Local Similarity 48.0%; Pred. No. 1.9; Indels 0; Gaps 0;
 Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 374 CACCTCGGCCATGGACAGTGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
 Db 324 CTCCCTCGTGTGGCTGGCATGACACCCGGATGATGTTGGTGGATGAC 265

Qy 434 GCCAGTGTCTGAGCTGGGGCAACTTCACCTGAGAACAGGGCAGGCCAGGGT 493
 Db 264 GGCGAGGTGGAGTTGGGATGATCCTCGATGAGGGCACTGGTCTCGTAGGG 205

Qy 494 GGAGATGATCTGCAGGGTCCGGGAGGCCACCTATACCAACAGCCATGCGGA 553
 Db 204 CTGTATCACCACTATGCGCCCTGGGAGCTGTGATGTCGAGTCGATCGGGT 145

Qy 554 GGATGGCAGGTGCACTTGAGAGAACATGCCACAGGCG 596
 Db 144 GGAGGCCGCGTAGTGTGATGTCGATGGTGAACATGGCG 102

RESULT 25
 US-09-095-855-202/C
 Sequence 202, Application US/09095855
 ; Patent No. 6160093

GENERAL INFORMATION:
 ; APPLICANT: Tan, Paul
 ; ATTORNEY: Visser, Elizabeth
 ; SKINNER, MARCOT
 ; APPLICANT: Prestidge, Ross
 ; TITLE OF INVENTION: Compounds and Methods for
 ; NUMBER OF SEQUENCES: 208
 ; CORRESPONDENCE ADDRESS:
 ; STREET: 2601 Elliott Avenue, Suite 4185
 ; CITY: Seattle
 ; STATE: WA
 ; COUNTRY: USA
 ; ZIP: 98121
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/095,855
 ; FILING DATE:

RESULT 24
 US-09-205-426-179/C
 Sequence 179, Application US/09205426

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/705,347
 FILING DATE: 29-AUG-1996
 APPLICATION NUMBER: 08/873,970
 FILING DATE: 12-JUN-1997
 APPLICATION NUMBER: 08/957,362
 FILING DATE: 23-DEC-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Sleath, Janet
 REGISTRATION NUMBER: 37,007
 REFERENCE/DOCKET NUMBER: 11000.10002c3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 206-269-0565
 TELEFAX: 206-269-0563
 TELEX:
 INVENTION FOR SEQ ID NO: 202:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 570 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-09-095-855-202

Query Match Score 37.4; DB 3; Length 570;
 Best Local Similarity 3.8%; Pred. No. 2;
 Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 374 CACCTCATGGCAGTGGACAGGTACAGTGCAAGGCTACTGGGAGCTGGTGCACAA 433
 Db 333 CTCTCGTGAATGCCGATTCGCGATACCCGGATGATGTTGCTGGATTGAC 274
 Qy 434 GCGAGTGCCTGAGCTGGCCACTCTGAGGAGGGCAGAGGGCCACGGT 493
 Db 273 GCGGAGGTCGAGCTGGCTCCCTGGAGGCCACCTATCACCAAGCTGATCGGGAA 553
 Qy 554 GGTGGGCGGGTCCACCTGCGAGGACATGCAAGGGCAGTGTGATGGGGT 596
 Db 153 GGAGGGCCGGTAGTAGTGTGATGTGATCGTGGATCGTGGATGGCG 111

RESULT 26
 US-09-205-426-202/c
 Sequence 202 Application US/09205426
 GENERAL INFORMATION:
 APPLICANT: Watson, James D.
 APPLICANT: Tan, Paul L. J.
 TITLE OF INVENTION: Compounds and Methods for Treatment and
 Diagnosis of Mycobacterial Infections
 FILE REFERENCE: 11000.1002c4
 CURRENT APPLICATION NUMBER: US/09/205,426
 FILING DATE: 1998-12-04
 EARLIER APPLICATION NUMBER: 09/095,855
 EARLIER FILING DATE: 1998-09-11
 EARLIER APPLICATION NUMBER: 08/997,362
 EARLIER FILING DATE: 1997-12-23
 EARLIER APPLICATION NUMBER: 08/873,970
 EARLIER FILING DATE: 1997-06-12
 EARLIER APPLICATION NUMBER: 08/705,347
 EARLIER FILING DATE: 1996-08-29
 NUMBER OF SEQ ID NOS: 208
 SOFTWARE: FastSEQ for Windows Version 3.0
 SEQ ID NO 202
 LENGTH: 570
 TYPE: DNA
 ORGANISM: Mycobacterium vaccae

Classification: 3.8%; Score 37.4; DB 4; Length 570;
 Best Local Similarity 48.0%; Pred. No. 2;
 Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 374 CACCTCATGGCAGTGGACAGGTACAGTGCAAGGCTACTGGGAGCTGGTGCACAA 433
 Db 333 CTCTCGTGAATGCCGATTCGCGATACCCGGATGATGTTGCTGGATTGAC 274
 Qy 434 GCGAGTGCCTGAGCTGGCCACTCTGAGGAGGGCAGAGGGCCACGGT 493
 Db 273 GCGGAGGTCGAGCTGGCTCCCTGGAGGCCACCTATCACCAAGCTGATCGGGAA 553
 Qy 554 GGTGGGCGGGTCCACCTGCGAGGACATGCAAGGGCAGTGTGATGGGGT 596
 Db 153 GGAGGGCCGGTAGTAGTGTGATGTGATCGTGGATCGTGGATGGCG 111

RESULT 27
 US-09-119-B9z-3
 Sequence 3 Application US/09319892
 GENERAL INFORMATION:
 PATENT NO. 6117616
 APPLICANT: BARTSCH, Klaus
 APPLICANT: KRIETE, Guido
 APPLICANT: BROER, Inge
 APPLICANT: PURLER, Alfred
 TITLE OF INVENTION: NOVEL GENES CODING FOR AMINO ACID DEACETYLASES WITH
 TITLE OF INVENTION: SPECIFICITY FOR N-ACETYL-L-PHOSPHINOTRICIN, THEIR
 TITLE OF INVENTION: ISOLATION AND THEIR USE
 FILE REFERENCE: 514412-2005
 CURRENT APPLICATION NUMBER: US/09/319,892
 CURRENT FILING DATE: 1999-06-14
 EARLIER APPLICATION NUMBER: PCT/EP97/06755
 EARLIER FILING DATE: 1997-12-03
 EARLIER APPLICATION NUMBER: 19652284-6
 EARLIER FILING DATE: 1996-12-16
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 3
 LENGTH: 1273
 TYPE: DNA
 ORGANISM: Stenotrophomonas maltophilia
 US-09-319-892-3

Query Match Score 36.6; DB 3; Length 1273;
 Best Local Similarity 45.0%; Pred. No. 4.3;
 Matches 138; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

Qy 431 CAAGCCAGTGTCTGAGCTGGGCCAACTTCTGAGGAGGGGAGGGCCAG 490
 Db 735 CGAGGCCACGAGCTGCTGCCGACAGGGTGAAGGCTGACCTTCAC 794
 Qy 491 CGTGGGAGATGATGCTGCGAGGGCTCTGGAGGCCACCTATCACCAAGCTGATCGGG 550
 Db 795 GCTGGGAGGTGCTGGATCGAGGCTGGCGAGACATCTGCG 854
 Qy 551 GAAGGATGGAGGTCCACCTGCGAGAAGACCATGCAAGGACCTCTGCAC 610
 Db 855 GGCGCATGACAAGCGCTGGAGTTGAGTGTGGCAACTACCCCCCACCATCCTACTC 914
 Qy 611 CTTCCCTGCGAGGCCAGACATGGACTGTTCTGGTCCAGGGCTGCAAACAGCCCATG 670
 Db 915 CGCCCGGAGGGCGGAGCTGGCTGAGGTGCTGAGCTGGGAGCCAAACAGCTG 974
 Qy 671 CCAGCAAGGCCCTCACTGGCTGCCAACAGGGCTGAGGACTGGCCA 730
 Db 975 GCTGCCCAAGGCCCTCACTGGCTGCCAACAGGGCTGAGGACTGGCCA 1034

Db 863 GC 864

RESULT 31
US-09-252-991A-11094/C
; Sequence 1.094, Application US/09252991A

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11094
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11094

Qy 939 AAAATGAGTGTCTTTAGCTGCTCTGCCACAAAAAA^{AAAAAAAA}
Db 1950 AAAACAGTGTITTAAGTCCGAAGTCCC^{AAAAAAAA}
US-09-340-993-12

RESULT 33
US-09-340-993-12
; Sequence 1.094, Application US/09340993

; GENERAL INFORMATION:
; Patent No. 63034228
; APPLICANT: Tyrell E. No. 6034228ris
; APPLICANT: William Craig Moore
; APPLICANT: David Shay Silberstein
; TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION SERINE/THREONINE KINASE
; FILE REFERENCE: PHM-70296_N1
; CURRENT APPLICATION NUMBER: US/09/340,993
; CURRENT FILING DATE: 1999-06-25
; EARLIER APPLICATION NUMBER: GB 9726851.0 & US 09/211,930
; EARLIER FILING DATE: 1997-12-19 & 1998-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Mus musculus

US-09-340-993-12

Query Match 3.7%; Score 36.4; DB 4; Length 1584;
Best Local Similarity 50.0%; Pred. No. 5.2; Indels 0; Gaps 0;
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 601 CCAACTCTCTTCCTCGCCAGCAGCATCGACTGGTCTGGGCCAGGTGAAACA 660
Db 839 CCAAGGCTTTCGCCGAGCGGCCCTCGACAGGCTGTGCCGGATCAGCTCAAGGAAGCAGGG 780

Qy 661 AGCCCATGTCAGCACAGGCCCTCACAGTGTGCCCGCAGGTGGTGAACAGAGATGG 720
Db 779 ACGGCCACATCTGGTCCCGCCTGGCTGCTGGCCGAAGGATGGATGG 720

Qy 721 AGCACTGGAGGTCCCTGGAGGCCCATCCCTGCCTGCCGCTCTAGGGCACCC 780
Db 719 TCGAGGCCGACCGACATGGCATCCCTCCGGGACGAGCCTGAAC^{ACGCCCT} 660

Qy 781 GC 782
Db 659 GC 658

RESULT 34
US-09-468-442-12
; Sequence 1.094, Application US/09468442

; GENERAL INFORMATION:
; Patent No. 6300098
; APPLICANT: Tyrell E. No. 6300098ris
; APPLICANT: William Craig Moore
; APPLICANT: David Shay Silberstein
; TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION SERINE/THREONINE KINASE
; FILE REFERENCE: PHM-70296_N1
; CURRENT APPLICATION NUMBER: US/09/468,442
; CURRENT FILING DATE: 1999-1-2-21
; EARLIER APPLICATION NUMBER: US 09/340,993
; EARLIER FILING DATE: 1999-06-25
; EARLIER APPLICATION NUMBER: GB 9726851.0
; EARLIER FILING DATE: 1997-12-19
; EARLIER APPLICATION NUMBER: US 09/211,930
; EARLIER FILING DATE: 1998-1-2-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Mus musculus

US-09-468-442-12

Query Match 3.7%; Score 36.4; DB 4; Length 2028;
Best Local Similarity 79.6%; Pred. No. 5.7; Mismatches 0; Indels 0; Gaps 0;

Qy 939 AAAATGAGTGTCTTTAGCTGCTCTGCCACAAAAAA^{AAAAAAAA}
Db 1950 AAAACAGTGTITTAAGTCCGAAGTCCC^{AAAAAAAA}
US-09-211-930-12

Query Match 3.7%; Score 36.4; DB 2; Length 2028;
Best Local Similarity 79.6%; Pred. No. 5.7; Mismatches 0; Indels 0; Gaps 0;

Qy 939 AAAATGAGTGTCTTTAGCTGCTCTGCCACAAAAAA^{AAAAAAAA}
Db 1950 AAAACAGTGTITTAAGTCCGAAGTCCC^{AAAAAAAA}
US-09-252-991A-5589/c

RESULT 35
US-09-252-991A-5589/c

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT OF BODY WEIGHT DISORDERS INCLUDING OBESITY

FILE REFERENCE: 7B53-136

CURRENT APPLICATION NUMBER: US/09/245, 041

CURRENT FILING DATE: 1999-02-05

EARLIER APPLICATION NUMBER: 6/093, 630

EARLIER FILING DATE: 1998-07-21

EARLIER APPLICATION NUMBER: 6/0104, 978

EARLIER FILING DATE: 1998-10-20

NUMBER OF SEQ ID NOS: 131

SOFTWARE: Fast-SEQ For Windows Version 3.0

LENGTH: 2625

TYPE: DNA

ORGANISM: Homo sapiens

US-09-245-041-18

Query Match 3.6%; Score 36; DB 3; Length 2625;

Best Local Similarity 67.1%; Prod. No. 7.9;

Matches 51; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 91.7 TGTTCGATTGGAGTTATGCAAAATGAGTGTGTTAGCTCTGCCACAAAAA 976

Db 2528 TGTTCCTACTGCTCTTGAAATAAGTGAACATCCTTGTCTGTAAANAAA 2587

Qy 977 AAAAAAAAGAAAAA 992

Db 2588 AAAAAAAAAAAA 2603

RESULT 39

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; ATTORNEY: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103, 840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 3.6%; Score 36; DB 3; Length 4403765;

Best Local Similarity 47.9%; Pred. No. 69;

Matches 145; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 239 CTATTCCCTTGTTGGAAACCAGAACATCAAGTGGCCAAGAGGTGCTGAGACCGA 298

Db 3773464 CGACGCCCGTATTGGCCAAACTACCGCAGGATGCGCAAGTGGGTGAGGAGC 3773523

Qy 299 GCGGGCTCTCCTCAACTCAAGTCAACACTAAGTCAGTCCAGA CCTGGTCACCTACTT 358

Db 3773524 GATCGCGCCPACAGCACCCGGCTGAGATCACCCGACCGGGTGTGGAGACSCCCAC 3773583

Qy 359 CTGCGGGCTCTCACCTAGGTGCCATGTTGACAGTCAAGATGCACTG 418

Db 3773584 GGCTTAAGCGGGCTGGCCGCTGCAGTGAATCAAATCCA GACGGGTTGGCGCG 3773643

Qy 419 GGAGTGGCCCAAGGGTGGAGATGATGGCCAGGGCCTCTCG 518

Db 3781521 GCCGACGCCAGGGTGGAGGGCCATCCGGTCGGACG 3781560

RESULT 40

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103, 840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 441529

; TYPE: DNA

; OTHER INFORMATION: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-1

Query Match 3.6%; Score 36; DB 3; Length 4403765;

Best Local Similarity 47.9%; Pred. No. 69;

Matches 145; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 239 CTATTCCCTTGTTGGAAACCAGAACATCAAGTGGCCAAGAGGTGCTGAGACCGA 298

Db 3773464 CGACGCCCGTATTGGCCAACTACCGCAGGATGCGCAAGTGGGTGAGGAGC 3773523

Qy 299 GCGGGCTCTCCTCAACTCAAGTCAACACTAAGTCAGTCCAGA CCTGGTCACCTACTT 358

Db 3773524 GATCGCGCCPACAGCACCCGGCTGAGATCACCCGACCGGGTGTGGAGACSCCCAC 3773583

Qy 359 CTGCGGGCTCTCACCTAGGTGCCATGTTGACAGTCAAGATGCACTG 418

Db 3773584 GGCTTAAGCGGGCTGGCCGCTGCAGTGAATCAAATCCA GACGGGTTGGCGCG 3773643

Qy 419 GGAGTGGCCCAAGGGTGGAGATGATGGCCAGGGCCTCTCG 518

Db 3781521 GCCGACGCCAGGGTGGAGGGCCATCCGGTCGGACG 3781560

RESULT 41

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: SCHUPP, Thomas M.

; APPLICANT: LIGON, James Joseph

; APPLICANT: BECK, Dwight Steven

; APPLICANT: RYALS, John Andrew

; APPLICANT: GAFFNEY, Thomas Deane

; APPLICANT: LAM, Stephen Ting

; APPLICANT: HAMMER, Phillip E.

; APPLICANT: URENS, Scott Joseph

; TITLE OF INVENTION: Genes for the synthesis of antipathogenic substances

; NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/258,261B
 FILING DATE: 08-JUN-1994
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/457,205
 FILING DATE: 01-JUN-1995
 APPLICATION NUMBER: 08/457,205
 FILING DATE: 01-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: CGC 1506/CRP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8614
 TELEFAX: 919-541-8649
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 28958 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-258-261B-6

Query Match 3.6%; Score 35.8; DB 1; Length 28958;

Best Local Similarity 52.3%; Pred. No. 21;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy	56	CGAGGCAATGGGAATCTCCGGGTGTTGCTGGCTGCCAGAGCTCTC	115
Db	4498	CAGGGCAATTGGGGGTGAGGGCTTGGCTGGCTGCCAGAGCTCTC	4439
Qy	116	CAAGGACAGGGAGGAAGAATTACCCCTGTCCATTGCTACAAGTCCTGGAACT	175
Db	4438	GAAGGACGGAGGAGGACGGATGACCGTTTGAGCGCATCGAGGAGGA	4379
Qy	176	TTCCTCAAAGGCGCTGGTGTCAATACC	206
Db	4378	GCATCCCGGCGCCCTGCCCCAGCCGAAAC	4348

Query Match 3.6%; Score 35.8; DB 1; Length 28958;
 Best Local Similarity 52.3%; Pred. No. 21;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 Qy 56 CGAGGCAATGGGAATCTCCGGGTGTTGCTGGCTGCCAGAGCTCTC 115
 Db 4498 CAGGGCAATTGGGGGTGAGGGCTTGGCTGGCTGCCAGAGCTCTC 4439
 Qy 116 CAAGGACAGGGAGGAAGAATTACCCCTGTCCATTGCTACAAGTCCTGGAACT 175
 Db 4438 GAAGGACGGAGGAGGACGGATGACCGTTTGAGCGCATCGAGGAGGA 4379
 Qy 176 TTCCCCAAAGGCGTGGTGTCAATACC 206
 Db 4378 GCATCCCGGCGCCCTGCCCCAGCCGAAAC 4348

RESULT 42
 US-08-456-837-6/c

Sequence 6, Application US/08456837
 Patent No. 564774
 GENERAL INFORMATION:
 APPLICANT: Schupp, Thomas
 APPLICANT: Ligon, James M.
 APPLICANT: Beck, James Joseph
 APPLICANT: Hill, Dwight Steven
 APPLICANT: Ryals, John Andrew
 APPLICANT: Gaffney, Thomas Deane
 APPLICANT: Lam, Stephen Ting
 APPLICANT: Hammer, Phillip E.
 APPLICANT: Uernes, Scott Joseph
 TITLE OF INVENTION: Genes for the synthesis of
 TITLE OF INVENTION: Antipathogenic substances
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation

RESULT 43
 US-08-457-342-6/c

Sequence 6, Application US/08457342
 Patent No. 562898
 GENERAL INFORMATION:
 APPLICANT: Schupp, Thomas
 APPLICANT: Ligon, James M.
 APPLICANT: Beck, James Joseph
 APPLICANT: Hill, Dwight Steven
 APPLICANT: Ryals, John Andrew
 APPLICANT: Gaffney, Thomas Deane
 APPLICANT: Lam, Stephen Ting
 APPLICANT: Hammer, Phillip E.
 APPLICANT: Uernes, Scott Joseph
 TITLE OF INVENTION: Genes for the synthesis of
 TITLE OF INVENTION: Antipathogenic substances
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/457,342
 FILING DATE: 01-JUN-1995
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/457,205
 FILING DATE: 01-JUN-1995
 APPLICATION NUMBER: 08/258,261
 FILING DATE: 08-Jun-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REFERENCE/DOCKET NUMBER: 36,129
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8614
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 28958 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-457-342-6

Query Match 3.6%; Score 35.8; DB 1; Length 28958;

Best Local Similarity 52.3%; Pred. No. 21;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 Qy 56 CGCAGGCATGGGCTCCCTGGCTGGTCTGGCTGGCGCAGCTTCCTC 115
 Db 4498 CAGGGCCATTGGCGGGTGGGGTGGCCATTGGCTGGCTGGCTGGCTGGCGCAGCCCTC 4439
 Qy 116 CAAGGCACGGAGAAATTACCCCTGGTCTCCATGGCTAACAAAGTCTGGAAAGT 175
 Db 4438 GAAGGAAGGGCAGGACGGATGACCCGTTRGGACGCCATGGAGGGCCCTGGAGGAGA 4379
 Qy 176 TTTCCCCAGAACGGCTGGTGCTCATAAC 206
 Db 4378 GCATCCGGGCGCCCTGGCCAGGCCAAC 4348

RESULT 44

US-08-457-646A-6/c
 Sequence 6, Application US/08457646A
 Patient No. 5679560

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas
 APPLICANT: Ligon, James M.
 APPLICANT: Beck, James Joseph
 APPLICANT: Hill, Dwight Steven
 APPLICANT: Rvals, John Andrew
 APPLICANT: Gaffney, Thomas Deane
 APPLICANT: Lam, Stephen Ting
 APPLICANT: Hammer, Phillip E.
 APPLICANT: Urnes, Scott Joseph
 TITLE OF INVENTION: Genes for the synthesis of
 TITLE OF INVENTION: antipathogenic substances
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532

COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/03/457,646A
 FILING DATE: 01-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/457,205
 FILING DATE: 01-JUN-1995
 APPLICATION NUMBER: 08/258,261
 FILING DATE: 08-Jun-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8614
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 28958 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-457-646A-6

Query Match 3.6%; Score 35.8; DB 1; Length 28958;

Best Local Similarity 52.3%; Pred. No. 21;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 Qy 56 CCGAGGCATGGGCTCCCTGGCTGGCTGGCGCAGCTTCCTC 115
 Db 4498 CAGGGCCATTGGCGGGTGGGGTGGCCATTGGCTGGCTGGCTGGCGCAGCCCTC 4439
 Qy 116 CAAGGCACGGAGAAATTACCCCTGGTCTCCATGGCTAACAAAGTCTGGAAAGT 175
 Db 4438 GAAGGAAGGGCAGGACGGATGACCCGTTRGGACGCCATGGAGGGCCCTGGAGGAGA 4379
 Qy 176 TTTCCCCAGAACGGCTGGTGCTCATAAC 206
 Db 4378 GCATCCGGGCGCCCTGGCCAGGCCAAC 4348

RESULT 45

US-08-458-076A-6/c
 Sequence 6, Application US/08458076A
 Patient No. 5698125

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas
 APPLICANT: Ligon, James M.
 APPLICANT: Beck, James Joseph
 APPLICANT: Hill, Dwight Steven
 APPLICANT: Rvals, John Andrew
 APPLICANT: Gaffney, Thomas Deane
 APPLICANT: Lam, Stephen Ting
 APPLICANT: Hammer, Phillip E.
 APPLICANT: Urnes, Scott Joseph
 TITLE OF INVENTION: Genes for the synthesis of
 TITLE OF INVENTION: antipathogenic substances
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/458,076A
 FILING DATE: 01-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/457,205
 FILING DATE: 01-JUN-1995
 APPLICATION NUMBER: 08/258,261
 FILING DATE: 08-Jun-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8614
 TELEFAX: 919-541-8689
 SEQUENCE CHARACTERISTICS:
 LENGTH: 28958 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-458-076A-6

Query Match 3.6%; Score 35.8; DB 1; Length 28958;
 Best Local Similarity 52.3%; Pred. No. 21;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy	56	ccggccatgggtccctgggtttcggttgcgttgtccatggcaggcgcttc	115
Db	4498	caggccattggccatggccgttgaggccatggccatggccatggccatggcgttc	4439
Qy	116	caaggacatggggggaaataccctgtggtcattgcctacaatggaaatgt	175
Db	4438	gaggacatggggggaaataccctgtggtcattgcctacaatggaaatgt	4379
Qy	176	tttccccaaaatggcggtttgggtcataac	206
Db	4378	gcatccccaaaatggcggtttgggtcataac	4348

RESULT 46
 US-08-458-335A-6/C
 Sequence 4, Application US/08764233A
 Patent No. 571649
 GENERAL INFORMATION:
 APPLICANT: Ligon, James M.
 APPLICANT: Schupp, Thomas
 APPLICANT: Beck, James J.
 APPLICANT: Hill, Dwight S.
 APPLICANT: Neff, Snezanna
 APPLICANT: Ryals, John A.
 TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 STREET: Ciba-Geigy Corporation
 CITY: Tarrytown
 STATE: NY
 COUNTRY: USA

ZIP: 10591
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/764,233A
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/258,261
 FILING DATE: 09-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/258,261
 FILING DATE: 08-JUN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: 1506/CIP6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 541-8587
 TELEFAX: (919) 541-8689
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 28958 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Sorangium cellulosum
 IMMEDIATE SOURCE:
 CLONE: P98/1
 US-08-764-233A-4

Query Match 3.6%; Score 35.8; DB 1; Length 28958;
 Best Local Similarity 52.3%; Pred. No. 21;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy	56	ccggccatgggtccctgggtttcggttgcgttgtccatggcaggcgcttc	115
Db	4498	caggccattggccatggccgttgaggccatggccatggccatggccatggcgttc	4439
Qy	116	caaggacatggggggaaataccctgtggtcattgcctacaatggaaatgt	175
Db	4438	gaggacatggggggaaataccctgtggtcattgcctacaatggaaatgt	4379
Qy	176	tttccccaaaatggcggtttgggtcataac	206
Db	4378	gcatccccaaaatggcggtttgggtcataac	4348

RESULT 47
 US-08-457-335A-6/C
 Sequence 6, Application US/08457335A
 Patent No. 5723759
 GENERAL INFORMATION:
 APPLICANT: Schupp, Thomas
 APPLICANT: Ligon, James M.
 APPLICANT: Beck, James J.
 APPLICANT: Hill, Dwight S.
 APPLICANT: Gaffney, Thomas Deane
 APPLICANT: Lam, Stephen Ting
 APPLICANT: Hammer, Phillip E.
 APPLICANT: Ryals, Scott Joseph
 TITLE OF INVENTION: Genes for the synthesis of
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation
 STREET: 7 Skyline Drive

CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/457,335A
 FILING DATE: 01-JUN-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/457,205
 FILING DATE: 01-JUN-1995
 APPLICATION NUMBER: 08/258,261
 FILING DATE: 08-Jun-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: CGC 1506/CIPS3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8614
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 28958 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-457-335A-6

RESULT 48
 Sequence 6, Application US/08729214
 Patent No. 5,817,502
 GENERAL INFORMATION:
 APPLICANT: Ligon, James M.
 APPLICANT: Hill, Dwight Steven
 APPLICANT: Ryals, John Andrew
 APPLICANT: Hammer, Phillip E.
 APPLICANT: van Pee, Karl-Heinz
 APPLICANT: Kirner, Sabine
 TITLE OF INVENTION: Genes for the synthesis of
 NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation
 STREET: 520 White Plains Road
 CITY: Tarrytown
 STATE: NY
 COUNTRY: USA

ZIP: 10591
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/729,214
 FILING DATE: TBA
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: CGC 1506/CIPS5
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 28958 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-729-214-6

Query Match 3.6%; Score 35.8; DB 1; Length 28958;
 Best Local Similarity 52.3%; Prod. No. 21;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 56 CGAGGCATGGCTCCCTGGCTGCTGTCTGCTGGCGAGCTGCTTC 115
 Db 4498 CAGGGCATTTGGGGTGAAGCCTGAGCCCTGCCTGTCAGGGCTC 4379

Qy 116 CAAGGACGGAGAAATTACCCCTGCTCATACANGTCTCTGAAAGT 175
 Db 4438 GAAGGAGGGAGGGATGACGGTGTGGTGTCAAAAC 206

Qy 56 CGGAGCATGGCTCCCTGGCTGCTCTGCTGGCGAGCTGCTTC 115
 Db 4498 CAGGGCATTTGGGGTGAAGCCTGAGCCCTGCCTGTCAGGGCTC 4348

RESULT 49
 Sequence 6, Application US/09028934
 GENERAL INFORMATION:
 APPLICANT: Ligon, James M.
 APPLICANT: Hill, Dwight S.
 APPLICANT: Lam, Steven T.
 APPLICANT: Hammer, Phillip E.
 APPLICANT: van Pee, Karl-Heinz
 APPLICANT: Kirner, Sabine
 APPLICANT: Young, Thomas R.
 APPLICANT: Hill, Dwight S.
 APPLICANT: Lam, Steven T.
 APPLICANT: Hammer, Phillip E.
 APPLICANT: van Pee, Karl-Heinz
 TITLE OF INVENTION: Pyrrolointron Biosynthesis Genes and Uses
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NO. 6112670artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709

Computer readable form:
 Medium type: Floppy disk
 Computer: IBM PC compatible
 Operating system: PC-DOS/MS-DOS
 Software: PatentIn Release #1.0, Version #1.30
 Current application data:

APPLICATION NUMBER: US/09/028,934
 FILING DATE:
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/729,214
 FILING DATE: 09-OCT-1996
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/258,261
 FILING DATE: 08-JUN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 8,241
 TELECOMMUNICATION INFORMATION:
 REFERENCE/DOCKET NUMBER: 1506/CIP26
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 541-8587
 TELEFAX: (919) 541-8689
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 49377 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Sorangium cellulosum
 IMMEDIATE SOURCE:
 CLONE: p98/1, pJL3, and pVKM15
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 383_760
 OTHER INFORMATION: /product= "SorR"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKS that are known to be involved in the synthesis of polyketide compounds."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 927_19874
 OTHER INFORMATION: /product= "Sora"
 OTHER INFORMATION: /note= "Gene Product is highly homologous to type I PKS such as eryA from Sacccharopolyspora erythraea."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 439_4439
 OTHER INFORMATION: /product= "Module 1 of SorA"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKS that are known to be involved in the synthesis of polyketide compounds."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 4438_4439
 OTHER INFORMATION: /product= "Module 1 of SorA"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKS that are known to be involved in the synthesis of polyketide compounds."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 4438_4439
 OTHER INFORMATION: /product= "Module 1 of SorA"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKS that are known to be involved in the synthesis of polyketide compounds."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 4439_4440
 OTHER INFORMATION: /product= "Module 2 of SorA"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKS that are known to be involved in the synthesis of polyketide compounds."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 13455_19616
 OTHER INFORMATION: /product= "Module 3 of SorA"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKS that are known to be involved in the synthesis of polyketide compounds."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 19871_46318
 OTHER INFORMATION: /product= "SorB"
 OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS genes"
 TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cliba-Geigy Corporation
 STREET: 520 White Plains Road, P.O. Box 2005
 CITY: Tarrytown
 STATE: NY
 ZIP: 10591
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/764,233A
 FILING DATE:
 CLASSIFICATION: 435

LOCATION: 40190..46318
 OTHER INFORMATION: /product= "Module 5 of SorB"
 FEATURE: misc feature
 NAME/KEY: misc feature
 LOCATION: 46811..47891
 OTHER INFORMATION: /product= "SorM"
 /note= "The protein encoded by the *sorM* gene is highly
 homologous to the methyltransferase from *Streptomyces*
hygroscopicus that is involved in the synthesis of the
 polyketide ramicin."
 OTHER INFORMATION: US-08-764-233A-1

Score 35.8; DB 1; Length 49377;
 Best Local Similarity 52.3%; Pred. No. 26;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 Qy 56 CCGAGGCATGGGGCTCCCTGGGCTTCGGTCTGGCTGGCGCTGCCAGCAGCTTC 115
 Db 20801 CAGGGCATGGGCTGGGCTGGGAGGCCCTGGCTGGCTGGCGCTGGCTGGTGA 20742
 Qy 116 CAAGGACGGAGGAATAATTACCCCTGGTCTCAATGGCTAACAGTCCTGAAAT 175
 Db 20741 CAAGGACGGAGGAACGGATGACCCGTTGGACAGGSCATCGAGGGCTGAA 20682
 Qy 176 TTTCCCAATAGCCGTGGCTATAACC 206
 Db 20681 GCATCCGGCCCTGGCCAGCCGAA 20651

Search completed: February 9, 2004, 15:01:36
 Job time : 177 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 12:59:37 ; Search time 328 Seconds
(without alignments)
8164.154 Million cell updates/sec

Title: US-09-990-726-222
Perfect score: 992
Sequence: 1 ggacgaggcggaaactagg.....aaaaaaa.....aaaaaaa 992

Scoring table: IDENTITY_NUC
GapOp 10_0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : N_Geneseq_15Jun03; *

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT; *
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT; *
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT; *
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT; *
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT; *
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT; *
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT; *
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT; *
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT; *
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT; *
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT; *
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT; *
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT; *
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT; *
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT; *
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT; *
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT; *
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT; *
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT; *
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT; *
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT; *
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT; *
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT; *
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT; *
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT; *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARTES

Result No.	Score	Query	Match	Length	DB ID	Description
1	992	100.0	992	21	AAC53373	Human PRO809 nucleic acid
2	992	100.0	992	22	AAFA1176	Novel human secreted protein
3	992	100.0	992	24	ABK31610	Novel human secreted protein
4	992	100.0	992	25	ABX80267	Novel human secreted protein
5	992	100.0	992	25	ABX80771	Novel human secreted protein
6	992	100.0	992	25	ABX81154	Novel human secreted protein
7	992	100.0	992	25	ABX90244	Novel human secreted protein
8	992	100.0	992	25	ABX77855	Novel human secreted protein

9	992	100.0	992	25	ABX79451	Human secreted/trypsinogen encoding human polypeptide
10	992	100.0	992	25	ABX4090	Human PRO polypeptide
11	992	100.0	992	25	ABX17054	Membrane-bound protein
12	991	99.9	991	21	AA265030	Human secreted protein
13	843	85.0	1047	22	AAH64779	Human EST-derived protein
14	790.2	79.7	935	22	AAH99131	Human secreted protein
15	773.6	78.0	1047	21	AAC74260	Human biallelic protein
16	773.6	78.0	1047	25	ABT16794	Human secreted protein
17	773.6	78.0	1047	25	ABZ66993	Human G-protein codon
18	501.6	50.6	502	25	ABZ36822	Human cDNA encoding novel nervous system protein
19	501.6	50.6	724	22	AAS34076	Human transducin gene
20	292.8	29.5	32220	22	ABA20618	Human breast cancer
C	21	149.6	15.1	150	19	AAK12008
C	22	149.6	15.1	150	19	AAK12009
C	23	47.2	4.8	1272	24	ABX97141
C	24	47	4.7	53226	25	ABQ76996
C	25	47	4.7	87350	18	AAL38336
C	26	44.4	4.5	636	23	AAS68336
C	27	44.4	4.4	249999	25	ABZ60229
C	28	43.4	4.4	560	24	ABT07673
C	29	43	4.3	509	24	ABT07672
C	30	42.8	4.3	143899	24	AAL38336
C	31	41.8	4.2	752	21	AAF21712
C	32	41.8	4.2	1417	21	AAC77293
C	33	41.8	4.2	2059	21	AAC77637
C	34	41.8	4.2	2672	25	ABX63699
C	35	41.8	4.2	3471	22	AAT15559
C	36	41.8	4.2	3541	24	ABU53715
C	37	41	4.1	38736	24	ABN97955
C	38	39.6	4.0	276	24	ABN22420
C	39	39.4	4.0	3285	22	AAAS9222
C	40	39.4	4.0	3285	24	ABA50891
C	41	39.4	4.0	3285	24	AAE50322
C	42	39.4	4.0	394	22	AAE98202
C	43	39.2	4.0	1639	21	AAE98202
C	44	38.8	3.9	2025	24	ABK33656
C	45	38.8	3.9	56737	24	ABR89898
C	46	38.8	3.9	185695	25	ABQ77405
C	47	38.4	3.9	294	23	ABV61098
C	48	38	3.8	349980	24	ABD81848
C	49	38	3.8	349980	23	ABV18588
C	50	37.8	3.8	353	23	AAI199683
C	51	37.8	3.8	4403765	22	AAI199683
C	52	37.8	3.8	4411529	22	AAI199682
C	53	37.6	3.8	375	23	AAI199683
C	54	37.6	3.8	861	24	ABA03348
C	55	37.6	3.8	1048	21	AAQ96232
C	56	37.6	3.8	1396	16	AAQ98498
C	57	37.6	3.8	1396	18	AAQT97324
C	58	37.6	3.8	1602	24	ABA90349
C	59	37.6	3.8	1626	25	ABT131957
C	60	37.6	3.8	3083	25	AAI51205
C	61	37.6	3.8	3537	25	AAI51404
C	62	37.6	3.8	80959	25	AAU21959
C	63	37.6	3.8	520	20	AAZ11382
C	64	37.4	3.8	697	23	ABU15733
C	65	37.4	3.8	520	24	ABU82288
C	66	37.4	3.8	570	20	AAZ11394
C	67	37.4	3.8	3675	21	AAF21797
C	68	37.2	3.8	296	22	AAU21959
C	69	37.2	3.8	450	19	AAV25967
C	70	37	3.7	697	23	AAV25967
C	71	36.8	3.7	512	24	ABU82279
C	72	36.8	3.7	545	24	ABU08343
C	73	36.8	3.7	784	23	ABU11520
C	74	36.8	3.7	2693	23	ABU08342
C	75	36.9	3.7	2784	23	AAV45110
C	77	36.6	3.7	1273	19	ABU08344
C	78	36.4	3.7	1571	24	ABL0445
C	79	36.4	3.7	381	20	AAV878083
C	80	36.4	3.7	381	20	AAV87989
C	81	36.4	3.7	745	19	ABU08345
C	81	36.4	3.7	745	24	ABQ92020

Human secreted/trypsinogen encoding human polypeptide
Human PRO polypeptide
Membrane-bound protein
Human secreted protein
Human EST-derived protein
Human secreted protein
Human biallelic protein
Human secreted protein
Human G-protein co-regulated protein
Human CDNA sequence
Genomic sequence e
Human breast and ovarian cancer
Human membrane spanning protein
Human retroviral sequence
Human ORFX protein
Human cDNA sequencing
Human membrane spanning protein
Human retroviral sequence
Human ORFX protein
Clone m103-4 encoded human prostate DNA
Human cDNA encoding novel human polyribonucleotide
Novel human polyribonucleotide
Human colon cancer cDNA
Human hypoxanthine-guanine phosphoribosyltransferase
Human polynucleotide
Human Polynucleotide kinase
Human encoding a mannosidase
Human 50 kDa dystransferrin
Human (50 kDa) dystransferrin
Human polynucleotide kinase
High-affinity phosphotransferase
High-affinity phosphotransfase
Human gene sequencing
Human secreted protein
Human secreted protein
Nucleotide sequence
M vacccine GV-41 codon
Nucleotide sequence
Human breast and ovarian cancer
Cysteine protease
Drosophila melanogaster
Human ovarian cancer
Drosophila melanogaster
Drosophila melanogaster
Comamonas sp. N-AC
Human polynucleotide
EST clone DB809-5
EST clone FR251
Clone AB809-5 iso1
Human polynucleotide

• 111

RECOUPMENT OF POLYNUCLEOTIC ACID FROM BACTERIAL SEQUENCES GIVEN IN THE PRESENT INVENTION.

CC encode human secreted proteins. The PRO nucleic acids, polypeptides, agonists and antagonists are useful for treating a PRO related disorder. CC The PRO polypeptides are useful for diagnosing tumours, especially lung CC cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or liver tumour. The PRO polypeptides are useful for stimulating the proliferation of, or gene expression, in pericyte cells, for stimulating the proliferation or differentiation of chondrocyte cells, for stimulating the release of tumour necrosis factor-alpha from human blood, CC for stimulating or inhibiting the proliferation of normal human dermal CC fibroblast cells. The PRO polypeptide may also be used as molecular weight markers and for tissue typing. The PRO nucleic acids have CC applications in molecular biology, including use as hybridisation probes, CC and in chromosome and gene mapping. ABK3363-ABK3367 represent human CC PRO protein coding sequences of the invention.

XX Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

Query Match 100.0%; Score 992; DB 24; Length 992;
Best Local Similarity 100.0%; Pred. No. 2.7e-23.7; ID: Mismatches 0; Indels 0; Gaps 0;

QY	1	GGCACGAGCAGGAACTAGGAGTTCTCACTGTGCCGAGAAGGCCCTACACCCACCGA	60	XX	781	GCCGCTCTGAGTGAAGAGGATGGTTGGGGTTTCAAGGATAGGAATGGGGGGTCAGAGGAC	840
Db	1	GGCACGAGCAGGAACTAGGAGTTCTCACTGTGCCGAGAAGGCCCTACACCCACCGA	60	XX	781	GCCGCTCTGAGTGAAGAGGATGGTTGGGGTTTCAAGGATAGGAATGGGGGGTCAGAGGAC	840
QY	61	GCATGGGCTCCCTGGCTGTTCTGCTTCATGGAGCTGCTGCCAGACGCTTCCTCAAAGG	120	XX	841	GCAAAGCAGCAGGAACTGTAAGATGCTCCAGGAAAGCTCCAGGAAAGCTCCAGGAA	900
Db	61	GCATGGGCTCCCTGGCTGTTCTGCTTCATGGAGCTGCTGCCAGACGCTTCCTCAAAGG	120	XX	841	GCAAAGCAGCAGGAACTGTAAGATGCTCCAGGAAAGCTCCAGGAAAGCTCCAGGAA	900
QY	121	CACGGGAGAAGAAATTACCCCTGGTTCATGGTCCATTAAGAGTCTGGAGATTTTCC	180	XX	901	GGCATAGCGTGTGACTTCTGTTGGATTTGGAGTCAATGCTGAAATGCTGTTAGCTSC	960
Db	121	CACGGGAGAAGAAATTACCCCTGGTTCATGGTCCATTAAGAGTCTGGAGATTTTCC	180	XX	901	GGCATAGCGTGTGACTTCTGTTGGATTTGGAGTCAATGCTGAAATGCTGTTAGCTSC	960
QY	181	CCAAAGGGCGCTGGGTGCTCATACCTCTGTGACCCCGAACCCGACATCACCT	240	XX	961	TCTTGCCACAAAAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAA	992
Db	181	CCAAAGGGCGCTGGGTGCTCATACCTCTGTGACCCCGAACCCGACATCACCT	240	XX	961	TCTTGCCACAAAAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAA	992
QY	241	ATTCCTCTGTGAAACCAAGAACTCACTCAAGTGTGCAAGGGTGTGAAGACCCAGAGC	300	OS	ABX80267	ABX80267 standard; DNA; 992 bp.	
Db	241	ATTCCTCTGTGAAACCAAGAACTCACTCAAGTGTGCAAGGGTGTGAAGACCCAGAGC	300	XX	ABX80267	ABX80267 standard; DNA; 992 bp.	
QY	301	CGGCTCTCTTCAACCTCAACTCAACTCAACTCAACTCAACTCAACTCAACTCT	360	PN	ABX80267	(first entry)	
Db	301	CGGCTCTCTTCAACCTCAACTCAACTCAACTCAACTCAACTCAACTCAACTCT	360	XX	ABX80267	(first entry)	
QY	361	GGCGGGCTCTCCACCTCACTGGCTCCATGTGACAGCTGGCAGGCTACAGTACCTGG	420	XX	DE	Novel human secreted or transmembrane protein PRO791 DNA.	
Db	361	GGCGGGCTCTCCACCTCACTGGCTCCATGTGACAGCTGGCAGGCTACAGTACCTGG	420	XX	XX		
QY	421	AAGTGTGCTCHAGGCAACTGCTGAGTGTGGCCGCCCACCTCTACTCTCTGAGACAGGG	480	XX	XX		
Db	421	AAGTGTGCTCHAGGCAACTGCTGAGTGTGGCCGCCCACCTCTACTCTCTGAGACAGGG	480	XX	PR	05-NOV-1997; 97W0-US20069.	
QY	481	CAGGCCCAAGGGTGGAGATGATCTGCAAGGGTGTCTCGGAGCCAACTTACCTCTGAGACAGGG	540	XX	PR	16-SEP-1998; 98W0-US19330.	
Db	481	CAGGCCCAAGGGTGGAGATGATCTGCAAGGGTGTCTCGGAGCCAACTTACCTCTGAGACAGGG	540	XX	PR	17-OCT-1998; 98W0-US19437.	
QY	541	CCAAACTCTCTGAGGATGGAGGATGGCAGGAGACATGGACTGTTCTGGCAGGCTGAAACA	600	XX	PR	07-OCT-1998; 98W0-US21141.	
Db	541	CCAAACTCTCTGAGGATGGCAGGAGACATGGACTGTTCTGGCAGGCTGAAACA	600	XX	PR	01-DEC-1998; 98W0-US25108.	
QY	601	CCAAACTCTCTGAGGATGGCAGGAGACATGGACTGTTCTGGCAGGCTGAAACA	660	XX	PR	05-JAN-1999; 99W0-US0016.	
Db	601	CCAAACTCTCTGAGGATGGCAGGAGACATGGACTGTTCTGGCAGGCTGAAACA	660	XX	PR	08-MAR-1999; 99W0-US0508.	
QY	661	ACGCCAATGTCAGCAGGCACTGGCTCACTGGCTGCCCAACTGGTGTGACGAGATGG	720	XX	PR	02-JUN-1999; 99W0-US12232.	
Db	661	ACGCCAATGTCAGCAGGCACTGGCTCACTGGCTGCCCAACTGGTGTGACGAGATGG	720	XX	PR	15-SEP-1999; 99W0-US21090.	
QY	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	15-SEP-1999; 99W0-US21547.		
Db	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	30-NOV-1999; 99W0-US28133.		
QY	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	01-DEC-1999; 99W0-US28301.		
Db	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	01-DEC-1999; 99W0-US04414.		
QY	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	24-FEB-2000; 2000W0-US04914.		
Db	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	24-FEB-2000; 2000W0-US05094.		
QY	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	02-MAR-2000; 2000W0-US0541.		
Db	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	10-MAR-2000; 2000W0-US06319.		
QY	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	15-MAR-2000; 2000W0-US06384.		
Db	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	20-MAR-2000; 2000W0-US0737.		
QY	721	AGGACTGGGAGGGCTCCCTGGAGAGGCCATCCCTTGCGCTCTACAGGGAGCACCC	780	PR	30-MAR-2000; 2000W0-US08439.		

Zhan	PPI	WPI;	QY	Query M
	XX	DR	Db	Best Lo
		P-PS	QY	Matches
PPT	Nove	link	Db	
and	are	in m	QY	
PPT	trea	poly	Db	
PPT	tre	usef	QY	
XXX	Clai	stim	Db	
PS	XX	PRO9	QY	
XX	CC	PRO1	Db	
	CC	usef	QY	
	CC	usef	Db	
	CC	immu	QY	
	CC	endo	Db	
	CC	mamm	QY	
	CC	PRO1	Db	
	CC	stim	QY	
	CC	immu	Db	
	CC	reti	QY	
	CC	rod	Db	
	CC	diso	QY	
	CC	and	Db	
	CC	and	QY	
	CC	decr	Db	
	CC	neph	QY	
	CC	dise	Db	
	CC	prol	QY	
	CC	are	Db	
	CC	sequ	QY	
	XXX	SQ	Sequ	
		SQ		

PI Zhang Z;

XX WPI; 2003-247083/24.

XX P-DR P-FSDB; ABUS9108.

XX Novel isolated PRO polypeptides e.g., PROB26, PRO1068, PRO1184, PRO1346 and PRO1375, which stimulate proliferation of stimulated T-lymphocytes are therapeutically useful for enhancing immune response and in cancer treatments -

XX Claim 2: Fig 152; 649pp; English.

The invention describes an isolated human PRO polypeptide. The PRO polypeptides are useful in detecting PRO polypeptides in a sample, in linking a bioactive molecule to a cell expressing a PRO polypeptide, and in modulating at least one biological activity of a cell expressing a PRO polypeptide. PRO1312 stimulates hypertrophy of neonatal heart and is thus useful for treating cardiac insufficiency disorders. PRO1154 and PRO1186 stimulate adrenal cortical endothelial growth, and PRO536, PRO943, PRO828, PRO826, PRO1068 or PRO555, PRO826, PRO819, PRO1126, PRO1360 and PRO1387 induce c-Fos in endothelial cells, and are thus useful for treating conditions or disorders where angiogenesis would be beneficial, e.g. wound healing and antagonist of this polypeptide are useful for treating cancerous tumours. PRO812 inhibits vascular endothelial growth factor (VEGF) stimulated proliferation of endothelial cells and is thus useful for inhibiting endothelial cell growth in mammals which would be beneficial in inhibiting tumour growth. PRO826, PRO1068, PRO1184, PRO1346 and PRO1132 stimulate proliferation of stimulated T-lymphocytes and are therapeutically useful for enhancing immune response. PRO828, PRO826, PRO1068 or PRO1132 enhance survival of retinal neurons cells (PRO1132 is also enhances survival/proliferation of rod photoreceptor cells) and therefore are useful for treating retinal disorders of injuries, e.g. retinitis pigmentosum, AMD, PRO819, PRO813 and PRO1066 induce proliferation of mammalian kidney mesangial cells, and therefore are useful for treating kidney disorders associated with decreased mesangial cell function such as Berger disease or other nephropathies associated with diabetes, hypertension or Crohn's disease. PRO1310, PRO844, PRO1312, PRO1192 and PRO1387 induce the proliferation and/or differentiation of chondrocytes in culture and are thus useful for treating sports injuries, and arthritis. This sequence represents a novel human PRO protein polynucleotide.

Query	Match	100.0%	Score	992;	DB	25;	Length	992;
	Best Local	Similarity	100.0%	Pred.	No.	2..7..236;	Mismatches	
	Matches	992;	Conservative	0;	Indels	0;	Gaps	0;
Qy	1	GGCAAGGCCAGGAACCTAGGAGTTCATGCCGAGCAGAGGCCCTAGACCCCGAG	60					
Ddb	1	GGCAAGGCCAGGAACCTAGGAGTTCATGCCGAGCAGAGGCCCTAGACCCCGAG	60					
Qy	61	GCTGGGCTCCCTGGCTGTCTGTGGCGTGTCTGTGGCGAGGTTCCTCAAGG	120					
Ddb	61	GCTGGGCTCCCTGGCTGTCTGTGGCGTGTCTGTGGCGAGGTTCCTCAAGG	120					
Qy	121	CACGGGAGGAAGAAATTACCCCTGGTGTCTCATGCCTCACAAAGTCCCTGAA	180					
Ddb	121	CACGGGAGGAAGAAATTACCCCTGGTGTCTCATGCCTCACAAAGTCCCTGAA	180					
Qy	181	CCAAAGGCCCTGGCTCTCATAACCTGCTGTGCCCTCCAGGCCAACCGCCCATACCT	240					
Ddb	181	CCAAAGGCCCTGGCTCTCATAACCTGCTGTGCCCTCCAGGCCAACCGCCCATACCT	240					
Qy	241	ATCCCTGTGTAACCAAGAACATCAAGGTGGCCAAAGGTGGTAAGACCCAGAGC	300					
Ddb	241	ATCCCTGTGTAACCAAGAACATCAAGGTGGCCAAAGGTGGTAAGACCCAGAGC	300					
Qy	301	CGCTCTTCAACCTCAACCTCAACTCAAGTCAGTCCAGACTGTCACCTACTCT	360					
Ddb	301	CGCTCTTCAACCTCAACCTCAACTCAAGTCAGTCCAGACTGTCACCTACTCT	360					
Qy	361	GCGGGCTCTCACCTCAAGTGCAAGTGCAGGGTACAGTGCACTGCTGGG	420					

Db	601	CCAACTTCTCTGCGAGCAAGATCGAACTGGTTCTGTGCAAGTCAGGACAAACA	660	
Qy	661	ACGCCAATGTCAGCAGGCCCTCACAGTGCGCCAGGGTGGTACAGAGAAATGG	720	06-JAN-2000; 2000WO-US00376.
Db	661	ACGCCAATGTCAGCAGGCCCTCACAGTGCGCCAGGGTGGTACAGAGAAATGG	720	PR 11-FEB-2000; 2000WO-US03565.
Qy	721	AGGACTGCAGGGTCCCTGGAGAGGCCATCTTGCCTTGGCTCTAACAGAGACCC	780	PR 18-FEB-2000; 2000WO-US04341.
Db	721	AGGACTGCAGGGTCCCTGGAGAGGCCATCTTGCCTTGGCTCTAACAGAGACCC	780	PR 22-FEB-2000; 2000WO-US04444.
Qy	781	GCGCTCTAAGTGAAGAGGTTGGGGTTCAAGGATAAGGATGGGAGGTCAGGAC	840	PR 24-FEB-2000; 2000WO-US04944.
Db	781	GCGCTCTAAGTGAAGAGGTTGGGGTTCAAGGATAAGGATGGGAGGTCAGGAC	840	PR 02-MAR-2000; 2000WO-US05004.
Qy	841	GCAGAGCAGGACCATGTCAGGATGGGAGGTCAGGACAGGAGCTGCA	900	PR 02-MAR-2000; 2000WO-US05811.
Db	841	GCAGAGCAGGACCATGTCAGGATGGGAGGTCAGGACAGGAGCTGCA	900	PR 10-MAR-2000; 2000WO-US06319.
Qy	901	GCGCATCAGCAGGACAGTGGTCTGATTGGGATCTGAAATGAGTGTGTTAGTGC	960	PR 15-MAR-2000; 2000WO-US06834.
Db	901	GCGCATCAGCAGGACAGTGGTCTGATTGGGATCTGAAATGAGTGTGTTAGTGC	960	PR 20-MAR-2000; 2000WO-US07377.
Qy	961	TCTTGCCACAAAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAAAC	992	PR 30-MAR-2000; 2000WO-US08439.
Db	961	TCTTGCCACAAAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAAAC	992	PR 15-MAY-2000; 2000WO-US13338.
Qy	XX	ABX81154 standard; DNA; 992 bp.	XX	PR 22-MAY-2000; 2000WO-US13705.
DB	AC	ABX81154;	AC	PR 22-MAY-2000; 2000WO-US14042.
XX	XX	22-APR-2003 (first entry)	XX	PR 30-MAY-2000; 2000WO-US14941.
DE	XX	Novel human secreted or transmembrane protein PRO791 DNA.	DE	PR 02-JUN-2000; 2000WO-US15244.
XX	XX		XX	PR 28-JUL-2000; 2000WO-US20710.
XX	XX		XX	PR 11-AUG-2000; 2000WO-US222031.
XX	XX		XX	PR 17-MAY-2000; 2000WO-US23552.
XX	XX		XX	PR 23-AUG-2000; 2000WO-US23352.
XX	XX		XX	PR 24-AUG-2000; 2000WO-US23338.
XX	XX		XX	PR 08-NOV-2000; 2000WO-US30932.
XX	XX		XX	PR 01-DEC-2000; 2000WO-US32678.
XX	XX		XX	PR 01-JUN-2001; 2001WO-US06540.
XX	XX		XX	PR 01-JUN-2001; 2001WO-US17800.
XX	XX		XX	PR 20-JUN-2001; 2001WO-US19632.
XX	XX		XX	PR 29-JUN-2001; 2001WO-US21066.
XX	XX		XX	PR 09-JUL-2001; 2001WO-US21735.
XX	XX		XX	PR 16-JUN-1997; 97US-049787P.
XX	XX		XX	PR 17-OCT-1997; 97US-06220P.
XX	XX		XX	PR 12-NOV-1997; 97US-065186P.
XX	XX		XX	PR 13-NOV-1997; 97US-065311P.
XX	XX		XX	PR 24-NOV-1997; 97US-06770P.
XX	XX		XX	PR 25-FEB-1998; 98US-075945P.
XX	XX		XX	PR 20-MAR-1998; 98US-078910P.
XX	XX		XX	PR 28-APR-1998; 98US-08332P.
XX	XX		XX	PR 07-MAY-1998; 98US-084600P.
XX	XX		XX	PR 28-MAY-1998; 98US-087106P.
XX	XX		XX	PR 02-JUN-1998; 98US-087607P.
XX	XX		XX	PR 02-JUN-1998; 98US-087609P.
XX	XX		XX	PR 03-JUN-1998; 98US-087782P.
XX	XX		XX	PR 04-JUN-1998; 98US-08801P.
XX	XX		XX	PR 04-JUN-1998; 98US-088025P.
XX	XX		XX	PR 04-JUN-1998; 98US-088026P.
XX	XX		XX	PR 04-JUN-1998; 98US-088028P.
XX	XX		XX	PR 04-JUN-1998; 98US-088029P.
XX	XX		XX	PR 04-JUN-1998; 98US-088030P.
XX	XX		XX	PR 04-JUN-1998; 98US-088033P.
OS	OS	Homo sapiens.	OS	PR 04-JUN-1998; 98US-088226P.
XX	XX	US2003027985-A1.	XX	PR 05-JUN-1998; 98US-088467P.
PN	PN	06-FEB-2003.	PN	PR 05-JUN-1998; 98US-088212P.
XX	XX	06-FEB-2003.	XX	PR 05-JUN-1998; 98US-088217P.
PD	PD	06-FEB-2003.	PD	PR 09-JUN-1998; 98US-08855P.
XX	XX	06-FEB-2003.	XX	PR 10-JUN-1998; 98US-088734P.
PF	PF	14-NOV-2001; 2001US-0990562.	PF	PR 10-JUN-1998; 98US-088738P.
XX	XX	14-NOV-2001; 2001US-0990562.	XX	PR 11-JUN-1998; 98US-088816P.
PR	PR	05-NOV-1997; 97WO-US20069.	PR	PR 12-JUN-1998; 98US-088810P.
PR	PR	16-SEP-1997; 98WO-US19330.	PR	PR 16-JUN-1998; 98US-088824P.
PR	PR	17-SEP-1997; 98WO-US19437.	PR	PR 10-JUN-1998; 98US-088826P.
PR	PR	07-OCT-1997; 98WO-US21141.	PR	PR 11-JUN-1998; 98US-088858P.
PR	PR	01-DEC-1997; 98WO-US25108.	PR	PR 16-JUN-1998; 98US-08914P.
PR	PR	05-JAN-1998; 99WO-US00106.	PR	PR 11-JUN-1998; 98US-089532P.
PR	PR	08-MAR-1998; 99WO-US05088.	PR	PR 17-JUN-1998; 98US-089338P.
PR	PR	02-JUN-1998; 99WO-US12232.	PR	PR 17-JUN-1998; 98US-089598P.
PR	PR	15-SEP-1998; 99WO-US21050.	PR	PR 17-JUN-1998; 98US-089600P.
PR	PR	15-SEP-1998; 99WO-US21547.	PR	PR 17-JUN-1998; 98US-089653P.
PR	PR	30-NOV-1998; 99WO-US28313.	PR	PR 17-JUN-1998; 98US-089653P.
PR	PR	01-DEC-1998; 99WO-US28301.	PR	
PR	PR	16-DEC-1998; 99WO-US28634.	PR	
PR	PR	16-DEC-1998; 99WO-US30055.	PR	
PR	PR	20-DEC-1998; 99WO-US30911.	PR	
PR	PR	05-JAN-2000; 2000WO-US00219.	PR	

PR	20-AUG-1998;	98US-097214P.
PR	18-JUN-1998;	98US-089907P.
PR	18-JUN-1998;	98US-089908P.
PR	19-JUN-1998;	98US-089907P.
PR	19-JUN-1998;	98US-089908P.
PR	19-JUN-1998;	98US-089951P.
PR	22-JUN-1998;	98US-090246P.
PR	22-JUN-1998;	98US-090252P.
PR	22-JUN-1998;	98US-090254P.
PR	23-JUN-1998;	98US-090349P.
PR	23-JUN-1998;	98US-090355P.
PR	24-JUN-1998;	98US-090429P.
PR	24-JUN-1998;	98US-090431P.
PR	24-JUN-1998;	98US-090445P.
PR	24-JUN-1998;	98US-090446P.
PR	24-JUN-1998;	98US-090445P.
PR	24-JUN-1998;	98US-090472P.
PR	24-JUN-1998;	98US-090535P.
PR	24-JUN-1998;	98US-090639P.
PR	25-JUN-1998;	98US-090540P.
PR	25-JUN-1998;	98US-090542P.
PR	24-JUN-1998;	98US-090557P.
PR	25-JUN-1998;	98US-090676P.
PR	25-JUN-1998;	98US-090678P.
PR	25-JUN-1998;	98US-090690P.
PR	25-JUN-1998;	98US-090694P.
PR	25-JUN-1998;	98US-090695P.
PR	26-JUN-1998;	98US-090656P.
PR	26-JUN-1998;	98US-090862P.
PR	26-JUN-1998;	98US-090863P.
PR	01-JUL-1998;	98US-091360P.
PR	01-JUL-1998;	98US-091544P.
PR	02-JUL-1998;	98US-091478P.
PR	02-JUL-1998;	98US-091519P.
PR	02-JUL-1998;	98US-091626P.
PR	02-JUL-1998;	98US-091628P.
PR	02-JUL-1998;	98US-091632P.
PR	02-JUL-1998;	98US-091646P.
PR	02-JUL-1998;	98US-091673P.
PR	07-JUL-1998;	98US-091978P.
PR	07-JUL-1998;	98US-091982P.
PR	09-JUL-1998;	98US-092182P.
PR	10-JUL-1998;	98US-092412P.
PR	20-JUL-1998;	98US-093339P.
PR	30-JUL-1998;	98US-094651P.
PR	04-AUG-1998;	98US-095282P.
PR	04-AUG-1998;	98US-095385P.
PR	04-AUG-1998;	98US-095301P.
PR	04-AUG-1998;	98US-095302P.
PR	04-AUG-1998;	98US-095318P.
PR	04-AUG-1998;	98US-095321P.
PR	04-AUG-1998;	98US-095916P.
PR	10-AUG-1998;	98US-095929P.
PR	10-AUG-1998;	98US-096766P.
PR	17-AUG-1998;	98US-096112P.
PR	17-AUG-1998;	98US-096733P.
PR	17-AUG-1998;	98US-096731P.
PR	17-AUG-1998;	98US-096857P.
PR	17-AUG-1998;	98US-096891P.
PR	17-AUG-1998;	98US-096894P.
PR	17-AUG-1998;	98US-096895P.
PR	17-AUG-1998;	98US-096897P.
PR	17-AUG-1998;	98US-096949P.
PR	18-AUG-1998;	98US-096950P.
PR	18-AUG-1998;	98US-096951P.
PR	18-AUG-1998;	98US-097022P.
PR	19-AUG-1998;	98US-097141P.

841	GCAAGCAGGCTATGTAATGAACTTCCAGAGGCCAAGGACGGCGAGGACTGCA	900
901	GGCCATCAGGTGCACTGTTGATTGGAGTTATGCAAATGTTAGGTGTTAGCTGC	960
901	GGCCATCAGGTGCACTGTTGATTGGAGTTATGCAAATGTTAGGTGTTAGCTGC	960
961	TCTTGCCACAAAAAaaaaaaaaaaaaaaa	992
961	TCTTGCCACAAAAAaaaaaaaaaaaaaaa	992

ABX90244 standard; cDNA; 992 BP.

human secreted/transmembrane protein cDNA, #90.
fumuman; gene; ss; PRO; secreted; transmembrane; signal peptide;
pharmaceutical; diagnostic; therapeutic; gene therapy
lomo saniens.

JS2002160384-A1

卷之三

卷之三

6-SEP-1998; 98WO-US193330.

17-OCT-1998 : 98M0-11S21141

15-JAN-1999; 99W0-US00106
2225 2225; 99W0-US22108;

990 GOMO=HOMO=1000

99WO-US21090.

0-NOV-1999; 99WO-US28313.

1-DEC-1999; 99MO-US28634.

0-DEC-1999 ; 99WO-US30911.

6=JAN-2000: 2000W0=11S00376

1-FEB-2000; 2000WU=USU3365.

2-FEB-2000; 2000W0-US04414.

4-FEB-2000; 2000WO-US05004.

0-MAR-2000; 2000WO-US06319.

0-MAR-2000; 2000WO-US07377:

5-MAY-2000 : 20000W0-US13358

1-MAY-2000; 2000M0-US13/03;

U-MA1-2000; U-2000WU=US14941.

8-JUL-2000; 2000WO-US20710.

3-AUG-2000; 2000WO-US23522.

8-NOV-2000; 2000W0-US30952.

PR 20-JUN-2001; 2001IW-US19692; 2001IW-US210656
 PR 29-JUN-2001; 2001IW-US210656
 PR 09-JUL-2001; 2001IW-US1735;
 PR 16-JUN-1997; 97US-049787P
 PR 17-OCT-1997; 97US-0862250P
 PR 12-NOV-1997; 97US-06516P
 PR 13-NOV-1997; 97US-055311P
 PR 24-NOV-1997; 97US-066770P
 PR 25-FEB-1998; 98US-05745P
 PR 20-MAR-1998; 98US-078910P
 PR 28-APR-1998; 98US-0833322P
 PR 07-MAY-1998; 98US-08400P
 PR 07-MAY-1998; 98US-087010P
 PR 02-JUN-1998; 98US-087607P
 PR 02-JUN-1998; 98US-087059P
 PR 02-JUN-1998; 98US-087759P
 PR 03-JUN-1998; 98US-0878227P
 PR 04-JUN-1998; 98US-088021P
 PR 04-JUN-1998; 98US-088023P
 PR 04-JUN-1998; 98US-088026P
 PR 04-JUN-1998; 98US-088028P
 PR 04-JUN-1998; 98US-088029P
 PR 04-JUN-1998; 98US-088030P
 PR 04-JUN-1998; 98US-088033P
 PR 04-JUN-1998; 98US-088034P
 PR 05-JUN-1998; 98US-088167P
 PR 05-JUN-1998; 98US-088204P
 PR 05-JUN-1998; 98US-088212P
 PR 05-JUN-1998; 98US-088217P
 PR 05-JUN-1998; 98US-088322P
 PR 10-JUN-1998; 98US-088324P
 PR 10-JUN-1998; 98US-088734P
 PR 10-JUN-1998; 98US-088738P
 PR 10-JUN-1998; 98US-088742P
 PR 10-JUN-1998; 98US-088810P
 PR 10-JUN-1998; 98US-088824P
 PR 10-JUN-1998; 98US-088828P
 PR 11-JUN-1998; 98US-088838P
 PR 11-JUN-1998; 98US-088861P
 PR 11-JUN-1998; 98US-088865P
 PR 12-JUN-1998; 98US-0889105P
 PR 16-JUN-1998; 98US-09440P
 PR 16-JUN-1998; 98US-088812P
 PR 16-JUN-1998; 98US-089514P
 PR 17-JUN-1998; 98US-089532P
 PR 17-JUN-1998; 98US-089538P
 PR 17-JUN-1998; 98US-089588P
 PR 17-JUN-1998; 98US-089599P
 PR 17-JUN-1998; 98US-089600P
 PR 17-JUN-1998; 98US-089633P
 PR 18-JUN-1998; 98US-089801P
 PR 18-JUN-1998; 98US-089907P
 PR 18-JUN-1998; 98US-089910P
 PR 28-AUG-2001; 2001US-0941992.
 (GETH) GENENTECH INC.
 Ashkenazi AJ, Baker KP, Bot
 Ferrara N, Fong S, Gerber H
 Grimaldi JC, Gurney AL, Kl
 Roy MA, Stewart TA,
 Zhang Z,
 WPI; 2003-288106/28.
 P-PSDB; ABU60538.

Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ, Grimaldi JC, Gurney AL, Kijaviv IU, Napier MA, Pan J, Paozi NF, Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Zhang Z; *GRANT & DOCS INC.*

New transmembrane Polypeptides and nucleic acids encoding the Polypeptides, useful in gene therapy, in chromosome identification, as chromosome markers, or in generating probes -

Claim 2; Fig 150; 650pp; English.

The invention discloses isolated PRO secreted/transmembrane polypeptides comprising a sequence without signal peptide and the nucleic acid

encoding them. The polypeptides can be used to raise antibodies that specifically bind to the PRO polypeptide, for linking a biactive molecule to a cell expressing a PRO protein and for modulating at least one biological activity of a cell. The PRO polypeptides or polynucleotides are also useful in gene therapy, in chromosome identification, as chromosomal markers, or in generating probes. The PRO polypeptides are useful as molecular markers for protein electrophoresis, and the isolated nucleic acids may be used for recombinant expressing those markers. The PRO polypeptides and nucleic acids may also be used in tissue typing. Anti-PRO antibodies are useful in diagnostic assays for PRO, and in affinity purification of PRO from recombinant cell culture or natural sources. The sequences presented in ABX9083-ABX9068 are the genes encoding the primers amplifying and probes detecting the PRO polynucleotides of the invention.

Note: The sequence data for this patent is also available in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

Query Match 100 0%; Score 992; DB 25; Length 992;

Best Local Similarity 100.0%; Pred. No. 2.7e-236; Length 992;

Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GCGACAGCAGGAACTAGGAGGTCTACTGGCCGAGGAGGGCTAACCCACCGAG 60	Db	1	GGCACAGCAGGAACTAGGAGGTCTACTGGCCGAGGAGGGCTAACCCACCGAG 60	QY	61	GCATGGGGTCCCTGGCTTCTGGCTGGCTGGCTGGCTGGCTGGCTCAAGG 120	Db	61	GCATGGGGTCCCTGGCTTCTGGCTGGCTGGCTGGCTGGCTGGCTCAAGG 120
QY	121	CACGGGAGGAGAAATTACCCCTGTTGTCATGCCPACAAGTCCNGGAGTTTC 180	Db	121	CACGGGAGGAGAAATTACCCCTGTTGTCATGCCPACAAGTCCNGGAGTTTC 180	QY	181	CCAAGGCCCTGGTGGCTATAAATCTGCTGTCATGCCAACCCCTCATCT 240	Db	181	CCAAGGCCCTGGTGGCTATAAATCTGCTGTCATGCCAACCCCTCATCT 240
QY	241	ATTCCTCTGTGAAACATCAAGGGCCAAAGAGGTGGTGACCCACGCCATGAGC 300	Db	241	ATTCCTCTGTGAAACATCAAGGGCCAAAGAGGTGGTGACCCACGCCATGAGC 300	QY	301	CGGCCCTCTAACCTCAACCTCAAGGTCAACTCAAGTCAGTCAGCTGCACTCT 360	Db	301	CGGCCCTCTAACCTCAACCTCAAGGTCAACTCAAGTCAGTCAGCTGCACTCT 360
QY	361	GCGGGCGTCTCCACCTGGTGGCCATGTGACAGGCTACAGATGCACTGG 420	Db	361	GCGGGCGTCTCCACCTGGTGGCCATGTGACAGGCTACAGATGCACTGG 420	QY	481	CAGGCCCAAGGTGGAGATACTGCAAGCTGGCTCTGGCAAGCCACCTATCACCACA 540	Db	481	CAGGCCCAAGGTGGAGATACTGCAAGCTGGCTCTGGCAAGCCACCTATCACCACA 540
QY	421	AGCTSTGGTCAAGCAAGTCTGAGCTGGGGCAACTCTACTCTGAGCAAGGGG 480	Db	421	AGCTSTGGTCAAGCAAGTCTGAGCTGGGGCAACTCTACTCTGAGCAAGGGG 480	QY	541	GCCTGATCGGGAGGATGGCAGGTCCACCTGAGACATGGGACTGGTGTGGCTG 600	Db	541	GCCTGATCGGGAGGATGGCAGGTCCACCTGAGACATGGGACTGGTGTGGCTG 600
QY	601	CCAACCTTCCTCTGGCCAGCAAGCATGGGACTGGTGTGGCTG 660	Db	601	CCAACCTTCCTCTGGCCAGCAAGCATGGGACTGGTGTGGCTG 660	QY	661	AGGCCATGTCAAGCAAGGGCCCTAACAGTGTGCCCCAGGTGGTGGACAGAATGG 720	Db	661	AGGCCATGTCAAGCAAGGGCCCTAACAGTGTGCCCCAGGTGGTGGACAGAATGG 720
QY	721	AGGACTGGCAGGGTCCCCTGGAGAGCCCTACCTTGCCTGCCCCCTAACAGGAGCACCC 780	Db	721	AGGACTGGCAGGGTCCCCTGGAGAGCCCTACCTTGCCTGCCCCCTAACAGGAGCACCC 780	QY	781	GCCGCTGACTGAGTGGAGGGTCTGGGGCTTCTGGGGATGGGGTCTGGGGAC 840	Db	781	GCCGCTGACTGAGTGGAGGGTCTGGGGCTTCTGGGGATGGGGTCTGGGGAC 840
QY	841	GCAAAGCAGCAGCCATGTGAAATGAACTGGCCATGAGGCAAGGAGGACTGCA 900	Db	841	GCAAAGCAGCAGCCATGTGAAATGAACTGGCCATGAGGCAAGGAGGACTGCA 900	QY	901	GGCATCAGGGTGTGACTGTGTTGGATGTTGAGTGTGTTAGCTGC 960	Db	901	GGCATCAGGGTGTGACTGTGTTGGATGTTGAGTGTGTTAGCTGC 960
QY	961	TCTTGCACAAAAATGAGTGTGTTAGCTGTTAGTGC 992	Db	961	TCTTGCACAAAAATGAGTGTGTTAGCTGTTAGTGC 992	QY	992	TCCTGCACAAAAATGAGTGTGTTAGCTGTTAGTGC 992	Db	992	TCCTGCACAAAAATGAGTGTGTTAGCTGTTAGTGC 992

RESULT 8

ABX77855

ID ABX77855 standard; cDNA; 992 BP.

ABX77855;

XX DT 14-APR-2003 (first entry)

XX Human PRO polynucleotide #61.

XX XX Human; PRO; gene; ss; cytosolic; tumour; cancer; breast; lung; stomach; liver; horse; cow; dog; cat; sheep; pig; goat; rabbit; ADAPT; antibody-dependent enzyme mediated prodrug therapy.

XX OS Homo sapiens.

XX PN US2003027163-A1.

XX PD 06-FEB-2003.

XX XX XX 15-NOV-2001; 2001US-0997666.

XX PR 05-NOV-1997; 97WO-US20069.

PR 16-SEP-1998; 98WO-US199330.

PR 17-SEP-1998; 98WO-US199437.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US251018.

PR 05-JAN-1999; 99WO-US00106.

PR 08-MAR-1999; 99WO-US05028.

PR 02-JUN-1999; 99WO-US12252.

PR 15-SEP-1999; 99WO-US21090.

PR 15-SEP-1999; 99WO-US211547.

PR 30-NOV-1999; 99WO-US28313.

PR 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US2834.

PR 16-DEC-1999; 99WO-US30095.

PR 02-FEB-2000; 2000WO-US07377.

PR 24-FEB-2000; 2000WO-US0494.

PR 05-JAN-2000; 2000WO-US00212.

PR 06-JAN-2000; 2000WO-US00376.

PR 11-FEB-2000; 2000WO-US03565.

PR 15-MAR-2000; 2000WO-US04341.

PR 20-MAR-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US0494.

PR 02-MAR-2000; 2000WO-US05841.

PR 10-MAR-2000; 2000WO-US0619.

PR 15-MAR-2000; 2000WO-US06884.

PR 20-MAR-2000; 2000WO-US07377.

PR 30-MAY-2000; 2000WO-US08439.

PR 15-MAY-2000; 2000WO-US13359.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.

PR	28-JUL-2000;	2000WO-US20710.	98US-090472P.
PR	11-AUG-2000;	2000WO-US22031.	98US-090525P.
PR	23-AUG-2000;	2000WO-US23522.	98US-090540P.
PR	24-AUG-2000;	2000WO-US23328.	98US-090545P.
PR	08-NOV-2000;	2000WO-US30952.	98US-090557P.
PR	01-DEC-2000;	2000WO-US32678.	98US-090670P.
PR	28-FEB-2001;	2001WO-US06520.	98US-090675P.
PR	01-JUN-2001;	2001WO-US17800.	98US-090690P.
PR	29-JUN-2001;	2001WO-US19692.	98US-090694P.
PR	09-JUL-2001;	2001WO-US21066.	98US-090698P.
PR	16-JUN-1997;	97US-049787P.	98US-090699P.
PR	17-OCT-1997;	97US-052250P.	98US-090863P.
PR	12-NOV-1997;	97US-06516P.	98US-091136P.
PR	13-NOV-1997;	97US-065311P.	98US-091544P.
PR	24-NOV-1997;	97US-065770P.	98US-091478P.
PR	25-FEB-1998;	98US-075945P.	98US-091519P.
PR	20-MAR-1998;	98US-078910P.	98US-091620P.
PR	28-APR-1998;	98US-083322P.	98US-091622P.
PR	07-MAY-1998;	98US-084600P.	98US-091633P.
PR	28-MAY-1998;	98US-087106P.	98US-091646P.
PR	02-JUN-1998;	98US-087607P.	98US-091673P.
PR	02-JUN-1998;	98US-087609P.	98US-091798P.
PR	02-MAR-1998;	98US-087759P.	98US-091922P.
PR	03-JUN-1998;	98US-087827P.	98US-092182P.
PR	04-JUN-1998;	98US-088021P.	98US-092472P.
PR	04-JUN-1998;	98US-088025P.	98US-093319P.
PR	04-JUN-1998;	98US-088026P.	98US-094651P.
PR	04-JUN-1998;	98US-088028P.	98US-095222P.
PR	04-JUN-1998;	98US-088029P.	98US-095285P.
PR	04-JUN-1998;	98US-088030P.	98US-095311P.
PR	04-JUN-1998;	98US-088033P.	98US-095302P.
PR	04-JUN-1998;	98US-088326P.	98US-095318P.
PR	05-JUN-1998;	98US-0888167P.	98US-095421P.
PR	05-JUN-1998;	98US-088202P.	98US-095425P.
PR	05-JUN-1998;	98US-088212P.	98US-095916P.
PR	05-JUN-1998;	98US-088217P.	98US-095929P.
PR	09-JUN-1998;	98US-088655P.	98US-096012P.
PR	10-JUN-1998;	98US-088734P.	98US-096143P.
PR	10-JUN-1998;	98US-088738P.	98US-096146P.
PR	10-JUN-1998;	98US-088742P.	98US-096329P.
PR	10-JUN-1998;	98US-088810P.	98US-096757P.
PR	10-JUN-1998;	98US-088824P.	98US-096766P.
PR	10-JUN-1998;	98US-088828P.	98US-096768P.
PR	11-JUN-1998;	98US-088850P.	98US-096773P.
PR	11-JUN-1998;	98US-088861P.	98US-096791P.
PR	11-JUN-1998;	98US-088872P.	98US-096816P.
PR	12-JUN-1998;	98US-089105P.	98US-096819P.
PR	16-JUN-1998;	98US-089440P.	98US-096866P.
PR	16-JUN-1998;	98US-089512P.	98US-096895P.
PR	16-JUN-1998;	98US-089514P.	98US-097141P.
PR	17-JUN-1998;	98US-089532P.	98US-097218P.
PR	17-JUN-1998;	98US-089533P.	98US-097219P.
PR	17-JUN-1998;	98US-089538P.	98US-097610P.
PR	17-JUN-1998;	98US-089907P.	98US-097611P.
PR	18-JUN-1998;	98US-08998P.	98US-097612P.
PR	18-JUN-1998;	98US-08998P.	98US-097659P.
PR	19-JUN-1998;	98US-089941P.	98US-097952P.
PR	19-JUN-1998;	98US-089941P.	98US-097979P.
PR	19-JUN-1998;	98US-089960P.	98US-097986P.
PR	19-JUN-1998;	98US-089633P.	98US-097986P.
PR	18-JUN-1998;	98US-089801P.	98US-098014P.
PR	17-JUN-1998;	98US-090246P.	98US-098255P.
PR	22-JUN-1998;	98US-090242P.	98US-100634P.
PR	22-JUN-1998;	98US-090254P.	98US-100858P.
PR	13-JUN-1998;	98US-090343P.	98US-113296P.
PR	24-JUN-1998;	98US-090444P.	98US-123957P.
PR	24-JUN-1998;	98US-090445P.	

PR	23-JUN-1999;	99US-141037P.	
PR	07-JUL-1999;	99US-143049P.	
Query Match	Best Local Similarity 100.0% ; Matches 992; Conservative 0;	Score 992; DB 25; Length 992;	
Matches	Mismatches 0; Indels 0;	Pred. No. 2.7e-236;	
QY	1 GGACGAGCCAGGAACCTAGGAGGTTCATGGCCAGAGGSCCCTAACCCCACCGAG	60	Db 961 TCTTGGCACACAAAAAAAGAAAAAAA 992
Db	1 GCGCAGGCGCAGAAGACTGGAGTTCACTGGCTTCACTGGCCAGAAGGCTTAACCGAG	60	RESULT 9
QY	61 GATGGGGCTCCCTGGGTGTTCTGGCTGGCGTGTGGCCACCGAGTCTCAAGG	120	ID ABX79451 standard; cDNA; 992 BP.
Db	61 GATGGGGCTCCCTGGGTGTTCTGGCTGGCGTGTGGCCACCGAGTCTCAAGG	120	XX
QY	121 CACGGGAGGAGAAATTACCCCTGGTGTCTATAACCTGCTGAAAGTCTGGAAAGT	180	AC ABX79451;
Db	121 CACGGGAGGAGAAATTACCCCTGGTGTCTATAACCTGCTGAAAGTCTGGAAAGT	180	XX
QY	181 CCAAAGGCCGTGGTGCTATAACCTGCTGAAAGTCTGGAAAGGCCACACCCATCACCT	240	DT 17-APR-2003 (first entry)
Db	181 CCAAAGGCCGTGGTGCTATAACCTGCTGAAAGTCTGGAAAGGCCACACCCATCACCT	240	XX
QY	241 ATTCCCTCTGTGAAACCAAGAACATCAAGGGGCCAAAGAGTGTGAAAGACCACAGAGC	300	DB Human secreted/transmembrane protein cDNA, # 90.
Db	241 ATTCCCTCTGTGAAACCAAGAACATCAAGGGGCCAAAGAGTGTGAAAGACCACAGAGC	300	XX
QY	301 CGGCCCTCTAACCTCAAGCTCAACTCAAGCTCAACTCAAGCTCAACTCAAGCTCAACTCT	360	KW Human; gene; ss; PRO; secreted; transmembrane; signal peptide;
Db	301 CGGCCCTCTAACCTCAACCTCAACTCAAGCTCAACTCAAGCTCAACTCAAGCTCAACTCT	360	KW pharmaceutical; diagnostic; biosensor; bioreactor; tumour; therapeutic;
QY	361 GCGGGCGTCTCCACCTGGCCATGGCCAGCTGGCAGGGTACAGATGCACTGGG	420	KW colon cancer; lung cancer; breast cancer;cancer; gene therapy.
Db	361 GCGGGCGTCTCCACCTGGCCATGGCCAGCTGGCAGGGTACAGATGCACTGGG	420	OS Homo sapiens.
QY	421 AGCTGTGTCAAGCGAAGTGTCTGAGCTGGGGCAACATCACTCTGCAGGAGAGGGGG	480	XX
Db	421 AGCTGTGTCAAGCGAAGTGTCTGAGCTGGGGCAACATCACTCTGCAGGAGAGGGGG	480	PN US2002142961-A1.
QY	481 CAGGCCCGGGTGGAGATGATCTGGCCAGGTCTCTCGGGAGGCCACCTATCCAAAC	540	XX
Db	481 CAGGCCCGGGTGGAGATGATCTGGCCAGGTCTCTCGGGAGGCCACCTATCCAAAC	540	PD 03-OCT-2002.
QY	541 GCCTCATGGGAAGGATGGCAGGTCCACCTGGCAACATGGCAACAGGAGGCTG	600	XX
Db	541 GCCTCATGGGAAGGATGGCAGGTCCACCTGGCAACATGGCAACAGGAGGCTG	600	PF 19-NOV-2001; 2001US-0989721.
QY	601 CCAACTCTCTCTCTCTGGCCAGGCCACATGGCAACATGGCAACAGGAGGCTG	660	PR 05-NOV-1997; 97WO-US20069.
Db	601 CCAACTCTCTCTCTGGCCAGGCCACATGGCAACATGGCAACAGGAGGCTG	660	PR 17-SEP-1998; 98WO-US19437.
QY	661 AGCGCAATGTCAGGCCCTCACAGTGTGGCCAGGGTGTGCAAGGTGCAAAAC	720	PR 07-OCT-1998; 98WO-US21141.
Db	661 AGCGCAATGTCAGGCCCTCACAGTGTGGCCAGGGTGTGCAAGGTGCAAAAC	720	PR 01-DEC-1998; 98WO-US25108.
QY	721 AGGATGCGGGGGCCCCCTGGAGGCCCTCACAGTGTGGCCAGGGTGTGCAAGGATG	780	PR 05-JAN-1999; 99WO-US0106.
Db	721 AGGATGCGGGGGCCCCCTGGAGGCCCTCACAGTGTGGCCAGGGTGTGCAAGGATG	780	PR 08-JAR-1999; 99WO-US05028.
QY	781 GCCGCTGTGAGTGAGGAGGTTGGGGGTTCAAGGATGGAAATGGGAGGTCAAGGAC	840	PR 02-JUN-1999; 99WO-US21252.
Db	781 GCCGCTGTGAGTGAGGAGGTTGGGGGTTCAAGGATGGAAATGGGAGGTCAAGGAC	840	PR 15-SEP-1999; 99WO-US21090.
QY	841 GCAAGAGGAGGCAATGTAAGTGGAGGTCTGAGGATGGCTGAGGAGCTGCA	900	PR 15-SEP-1999; 99WO-US1547.
Db	841 GCAAGAGGAGGCAATGTAAGTGGAGGTCTGAGGATGGCTGAGGAGCTGCA	900	PR 30-NOV-1999; 99WO-US28313.
QY	901 GGCCATCGGGGTGCAATGGTGTGAAATGGGAGGTCAAGGATGGAAATGGGAGGTCA	960	PR 01-DEC-1999; 99WO-US28301.
Db	901 GGCCATCGGGGTGCAATGGTGTGAAATGGGAGGTCAAGGATGGAAATGGGAGGTCA	960	PR 01-DEC-1999; 99WO-US28634.
QY	961 TCTTGGCACACAAAAAAAGAAAAAAA 992		PR 16-DEC-1999; 99WO-US20095.
Db			PR 20-DEC-1999; 99WO-US20911.
			PR 05-FEB-2000; 2000WO-US0219.
			PR 06-JAN-2000; 2000WO-US0376.
			PR 11-FEB-2000; 2000WO-US03565.
			PR 18-FB-2000; 2000WO-US04341.
			PR 22-FEB-2000; 2000WO-US04414.
			PR 22-FEB-2000; 2000WO-US07377.
			PR 30-MAR-2000; 2000WO-US08439.
			PR 24-FEB-2000; 2000WO-US0504.
			PR 02-MAR-2000; 2000WO-US0376.
			PR 10-MAR-2000; 2000WO-US14042.
			PR 15-MAR-2000; 2000WO-US14941.
			PR 20-MAR-2000; 2000WO-US15264.
			PR 02-JUN-2000; 2000WO-US15264.
			PR 28-JUL-2000; 2000WO-US20710.
			PR 11-AUG-2000; 2000WO-US13358.
			PR 17-MAY-2000; 2000WO-US13705.
			PR 22-MAY-2000; 2000WO-US23522.
			PR 30-MAY-2000; 2000WO-US23329.
			PR 08-NOV-2000; 2000WO-US14941.
			PR 08-NOV-2000; 2000WO-US30952.
			PR 01-DEC-2000; 2000WO-US32678.
			PR 28-FEB-2001; 2001WO-US06520.
			PR 01-JUN-2001; 2001WO-US1780.
			PR 20-JUN-2001; 2001WO-US1962.
			PR 29-JUN-2001; 2001WO-US21056.
			PR 09-JUL-2001; 2001WO-US21745.
			PR 16-JUN-1997; 97WO-US21745.
			PR 28-FEB-2001; 2001WO-US06520.
			PR 01-JUN-1997; 97US-062250P.
			PR 12-NOV-1997; 97US-06180P.
			PR 13-NOV-1997; 97US-06531P.
			PR 24-NOV-1997; 97US-066770P.
			PR 25-FEB-1998; 98US-075945P.

QY	841	CGAAAGGAGCACCCATCTAGAATGAAACCGTCCAGAGCAGCAAGGGAGGACTGCA	900	PR	01-DEC-2000;	2000WO-US32678.
Db	841	GAAGAGGAGCACCCATCTAGAATGAAACCGTCCAGAGCAGCAAGGGAGGACTGCA	900	PR	28-FEB-2001;	2001WO-US05520.
QY	901	GCCCCATAGCGCAGCACTGTGTTGGAGTTCATGCCAAATGAGTGTGTTAGCTGC	960	PR	01-JUN-2001;	2001WO-US17800.
Db	901	GCCCCATAGCGCAGCACTGTGTTGGAGTTCATGCCAAATGAGTGTGTTAGCTGC	960	PR	20-JUN-2001;	2001WO-US1692.
QY	961	TCTTGCACAAAAAAAALAAAAAA	992	PR	29-JUN-2001;	2001WO-US21066.
Db	961	TCTTGCACAAAAAAAALAAAAAA	992	PR	09-JUL-2001;	2001WO-US21735.
QY	961	TCTTGCACAAAAAAAALAAAAAA	992	PR	16-JUN-1997;	97US-04978P.
Db	961	TCTTGCACAAAAAAAALAAAAAA	992	PR	17-OCT-1997;	97US-06223P.
XX				PR	12-NOV-1997;	97US-06518P.
XX				PR	13-NOV-1997;	97US-06531P.
XX				PR	24-NOV-1997;	97US-06677P.
XX				PR	25-FEB-1998;	98US-07545P.
XX				PR	20-MAR-1998;	98US-07831P.
XX				PR	28-AFR-1998;	98US-08312P.
XX				PR	07-MAY-1998;	98US-08460P.
XX				PR	28-MAY-1998;	98US-08710P.
XX				PR	02-JUN-1998;	98US-087607P.
AC	ABX64090;			PR	02-JUN-1998;	98US-08769P.
XX				PR	02-JUN-1998;	98US-08775P.
DT	26-FEB-2003	(First entry)		PR	03-JUN-1998;	98US-087827P.
XX				PR	04-JUN-1998;	98US-088021P.
DE	CDNA encoding human PRO809 polypeptide.			PR	04-JUN-1998;	98US-08812P.
XX				PR	04-JUN-1998;	98US-088026P.
KW	Human; PRO polypeptide; secreted protein; transmembrane protein;			PR	04-JUN-1998;	98US-088028P.
KW	Genetic disorder; autobacterial; immunosuppressive; transgenic;			PR	04-JUN-1998;	98US-088030P.
KW	Gene therapy; gene; ss.			PR	04-JUN-1998;	98US-088033P.
XX				PR	04-JUN-1998;	98US-088126P.
OS	Homo sapiens.			PR	05-JUN-1998;	98US-088167P.
XX				PR	05-JUN-1998;	98US-088202P.
PN	US2002103125-A1.			PR	05-JUN-1998;	98US-088212P.
XX				PR	05-JUN-1998;	98US-088217P.
PD	01-AUG-2002.			PR	09-JUN-1998;	98US-088655P.
XX				PR	10-JUN-1998;	98US-088734P.
PP	20-NOV-2001;	2001US-0989731.		PR	10-JUN-1998;	98US-088738P.
XX				PR	10-JUN-1998;	98US-088742P.
XX				PR	10-JUN-1998;	98US-088744P.
PR	05-NOV-1997;	97WO-US20069.		PR	10-JUN-1998;	98US-088826P.
PR	16-SEP-1998;	98WO-US19330.		PR	11-JUN-1998;	98US-088858P.
PR	17-SEP-1998;	98WO-US19437.		PR	11-JUN-1998;	98US-088861P.
PR	07-OCT-1998;	98WO-US221141.		PR	11-JUN-1998;	98US-088876P.
PR	01-DEC-1998;	98WO-US25108.		PR	12-JUN-1998;	98US-088880P.
PR	05-JAN-1999;	99WO-US00106.		PR	12-JUN-1998;	98US-088910P.
PR	08-MAR-1999;	99WO-US05028.		PR	16-JUN-1998;	98US-089410P.
PR	02-JUN-1999;	99WO-US12252.		PR	16-JUN-1998;	98US-089412P.
PR	15-SEP-1999;	99WO-US21190.		PR	16-JUN-1998;	98US-089511P.
PR	15-SEP-1999;	99WO-US21547.		PR	17-JUN-1998;	98US-089514P.
PR	30-NOV-1999;	99WO-US28113.		PR	18-JUN-1998;	98US-089516P.
PR	01-DEC-1999;	99WO-US25108.		PR	17-JUN-1998;	98US-089528P.
PR	01-DEC-1999;	99WO-US28313.		PR	17-JUN-1998;	98US-089538P.
PR	01-DEC-1999;	99WO-US28341.		PR	17-JUN-1998;	98US-089598P.
PR	01-DEC-1999;	99WO-US28334.		PR	17-JUN-1998;	98US-089599P.
PR	16-DEC-1999;	99WO-US30095.		PR	17-JUN-1998;	98US-089600P.
PR	20-DEC-1999;	99WO-US30111.		PR	17-JUN-1998;	98US-089633P.
PR	06-JAN-2000;	2000WO-US00419.		PR	18-JUN-1998;	98US-089801P.
PR	06-JAN-2000;	2000WO-US00376.		PR	18-JUN-1998;	98US-089907P.
PR	11-FEB-2000;	2000WO-US03365.		PR	18-JUN-1998;	98US-08998P.
PR	18-FEB-2000;	2000WO-US04311.		PR	28-AUG-2001;	2001US-094199Z.
PR	22-FEB-2000;	2000WO-US04414.		XX		
PR	24-FEB-2000;	2000WO-US04514.		PA	(GETH) GENENTECH LTD.	
PR	02-MAR-2000;	2000WO-US05004.		XX		
PR	10-MAR-2000;	2000WO-US05641.		PI	Ashkenazi AJ,	
PR	15-MAR-2000;	2000WO-US06319.		PI	Baker KP,	
PR	20-MAR-2000;	2000WO-US06834.		PI	Botstein D,	
PR	30-MAR-2000;	2000WO-US07377.		PI	Desnoyers L,	
PR	17-MAY-2000;	2000WO-US08419.		PI	Eaton DL,	
PR	17-MAY-2000;	2000WO-US13158.		PI	Ferrara N,	
PR	22-MAY-2000;	2000WO-US13705.		PI	Fgerber H,	
PR	30-MAY-2000;	2000WO-US14042.		PI	Gerritsen ME,	
PR	02-JUN-2000;	2000WO-US14941.		PI	Godowski PJ,	
PR	28-JUL-2000;	2000WO-US20764.		PI	Grimaldi JC,	
PR	11-AUG-2000;	2000WO-US20710.		PI	Gurney AL,	
PR	23-AUG-2000;	2000WO-US22031.		PI	Klajvin IU,	
PR	24-AUG-2000;	2000WO-US23322.		PI	Napier MA,	
PR	08-NOV-2000;	2000WO-US30952.		PI	Pan J,	
DR				PI	Paozi NF,	
DR				PI	Roy MA,	
DR				PI	Stewart TA,	
DR				PI	Williams PM,	
DR				PI	Zhang Z;	
WPI;				XX		
WPI;				XX		
P-PSDB;				XX		
ABU13920.				XX		
Claim 2; Fig 150; 64pp; English.				XX		

P-PSDB; ABU13920.

Novel secreted and transmembrane polypeptide for modulating biological activity of cell expressing the polypeptide, identifying agonists or antagonists of polypeptide, and as molecular weight markers -

XX

PS

Claim 2; Fig 150; 64pp; English.

Db	481	CAGGCCCAAGGGAGATCATGCGAGGCCCTCAGGGAGCCACTATCACCAACA	540
Qy	541	GCCATCGGGAAAGGATGGCAGGTACCTGCGAGAACATTGCAGGAGCTG	600
Db	541	GCCATCGGGAAAGGATGGCAGGTACCTGCGAGAACATTGCAGGAGCTG	600
Qy	601	CCAACCTCTCTCTCTGCGAGCGACATGGACGGTCTGTCAGCTGAAACA	660
Db	601	CCAACCTCTCTCTCTGCGAGCGACATGGACGGTCTGTCAGCTGAAACA	660
Qy	661	ACGCCAATGTCAGCACAGGCCCTCACAGTGCCCCAGGGTGAACAGATGG	720
Db	661	ACGCCAATGTCAGCACAGGCCCTCACAGTGCCCCAGGGTGAACAGATGG	720
Qy	721	AGGACTGGCAGGGTCCCTGGAGAGCCCATCCCTGCCCTACGGGACCC	780
Db	721	AGGACTGGCAGGGTCCCTGGAGAGCCCATCCCTGCCCTACGGGACCC	780
Qy	781	GCCGTCTGACTGAAGGGAGTTGGGGGTCAGGATGGAAATGGAGGGTCAAGGAC	840
Db	781	GCCGTCTGACTGAAGGGAGTTGGGGGTCAGGATGGAAATGGAGGGTCAAGGAC	840
Qy	841	GCAAGGAGCCAGGCTATGTAATGAACTGCCAGGAGCCAGGCTGCA	900
Db	841	GCAAGGAGCCAGGCTATGTAATGAACTGCCAGGAGCCAGGCTGCA	900
Qy	901	GGCCATCAGCGTGCAGCTGTGATTGGAGTTCTATGCAAATGAGTGTGTTTACGTGC	960
Db	901	GGCCATCAGCGTGCAGCTGTGATTGGAGTTCTATGCAAATGAGTGTGTTTACGTGC	960
Qy	961	TCTTGGCCAGAAAAAAAALAAAAAAALAAAAAA 992	
Db	961	TCTTGGCCAGAAAAAAAALAAAAAAALAAAAAA 992	
RESULT 12			
ID	AAZ65030	AAZ65030 standard; cdNA; 991 BP.	
XX	AAZ65030;		
AC			
XX			
DT	05-APR-2000	(first entry)	
XX			
DE	Membrane-bound protein PRO809 encoding cDNA.		
XX			
KW	Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;		
KW	pharmaceutical; receptor immunoadhesin; gene mapping; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W09963088-A2.		
XX			
PD	09-DEC-1999.		
XX			
PF	02-JUN-1999;	99WO-US12252.	
XX			
PR	02-JUN-1998;	98US-0087607.	
PR	02-JUN-1998;	98US-0087609.	
PR	02-JUN-1998;	98US-0087759.	
PR	03-JUN-1998;	98US-0087827.	
PR	04-JUN-1998;	98US-0088021.	
PR	04-JUN-1998;	98US-0088025.	
PR	04-JUN-1998;	98US-0088028.	
PR	04-JUN-1998;	98US-0088029.	
PR	04-JUN-1998;	98US-0088030.	
PR	04-JUN-1998;	98US-0088033.	
PR	04-JUN-1998;	98US-0088326.	
PR	05-JUN-1998;	98US-0088167.	
PR	05-JUN-1998;	98US-0088202.	
PR	05-JUN-1998;	98US-0088212.	
PR	09-JUN-1998;	98US-0088227.	
PR	10-JUN-1998;	98US-0088255.	
PR	10-JUN-1998;	98US-0088734.	
PR	10-JUN-1998;	98US-0088738.	
PR	10-JUN-1998;	98US-0088740.	
PR	10-JUN-1998;	98US-0088741.	
PR	10-JUN-1998;	98US-0088742.	
PR	10-JUN-1998;	98US-0088810.	
PR	10-JUN-1998;	98US-0088811.	
PR	10-JUN-1998;	98US-0088824.	
PR	10-JUN-1998;	98US-0088826.	
PR	10-JUN-1998;	98US-0088858.	
PR	11-JUN-1998;	98US-0088860.	
PR	11-JUN-1998;	98US-0088863.	
PR	11-JUN-1998;	98US-0088876.	
PR	12-JUN-1998;	98US-0089090.	
PR	12-JUN-1998;	98US-0089105.	
PR	16-JUN-1998;	98US-0089440.	
PR	16-JUN-1998;	98US-0089512.	
PR	16-JUN-1998;	98US-0089514.	
PR	17-JUN-1998;	98US-0089532.	
PR	17-JUN-1998;	98US-0089538.	
PR	17-JUN-1998;	98US-0089598.	
PR	17-JUN-1998;	98US-0089599.	
PR	17-JUN-1998;	98US-0089600.	
PR	17-JUN-1998;	98US-0089653.	
PR	18-JUN-1998;	98US-0089907.	
PR	18-JUN-1998;	98US-0089908.	
PR	18-JUN-1998;	98US-0089908.	
PR	19-JUN-1998;	98US-0089948.	
PR	19-JUN-1998;	98US-0089952.	
PR	22-JUN-1998;	98US-0090246.	
PR	22-JUN-1998;	98US-0090252.	
PR	23-JUN-1998;	98US-0090254.	
PR	23-JUN-1998;	98US-009349.	
PR	23-JUN-1998;	98US-0098355.	
PR	24-JUN-1998;	98US-0090429.	
PR	24-JUN-1998;	98US-009431.	
PR	24-JUN-1998;	98US-009445.	
PR	24-JUN-1998;	98US-00945.	
PR	24-JUN-1998;	98US-009461.	
PR	24-JUN-1998;	98US-009472.	
PR	24-JUN-1998;	98US-009535.	
PR	24-JUN-1998;	98US-00953.	
PR	24-JUN-1998;	98US-0095435.	
PR	24-JUN-1998;	98US-0095440.	
PR	24-JUN-1998;	98US-009557.	
PR	25-JUN-1998;	98US-009557.	
PR	24-JUN-1998;	98US-0095676.	
PR	25-JUN-1998;	98US-009678.	
PR	25-JUN-1998;	98US-009688.	
PR	25-JUN-1998;	98US-009690.	
PR	25-JUN-1998;	98US-009691.	
PR	25-JUN-1998;	98US-009694.	
PR	25-JUN-1998;	98US-009695.	
PR	25-JUN-1998;	98US-009696.	
PR	25-JUN-1998;	98US-009697.	
PR	26-JUN-1998;	98US-009698.	
PR	01-JUL-1998;	98US-0091358.	
PR	01-JUL-1998;	98US-0091360.	
PR	02-JUL-1998;	98US-0091544.	
PR	02-JUL-1998;	98US-0091478.	
PR	02-JUL-1998;	98US-0091519.	
PR	02-JUL-1998;	98US-0091626.	
PR	02-JUL-1998;	98US-0091628.	
PR	02-JUL-1998;	98US-0091633.	
PR	02-JUL-1998;	98US-0091646.	
PR	02-JUL-1998;	98US-0091673.	
PR	07-JUL-1998;	98US-009178.	
PR	07-JUL-1998;	98US-0091982.	
PR	09-JUL-1998;	98US-0091982.	
PR	10-JUL-1998;	98US-0092472.	

PR	20-JUL-1998;	98US-00933339.	XX	SQ Sequence 991 BP; 234 A; 301 C; 282 G; 174 T; 0 other;
PR	30-JUL-1998;	98US-0094451.	Query Match 99.9%; Score 991;	DB 21; Length 991;
PR	04-AUG-1998;	98US-00952825.	Best Local Similarity 100.0%; Pred. No. 4.be-236;	Mismatches 0; Indels 0; Gaps 0;
PR	04-AUG-1998;	98US-0095301.	Matches 991; Conservative 0;	
PR	04-AUG-1998;	98US-0095302.		
PR	04-AUG-1998;	98US-0095318.		
PR	04-AUG-1998;	98US-0095321.		
PR	04-AUG-1998;	98US-0095325.		
PR	10-AUG-1998;	98US-0095316.		
PR	10-AUG-1998;	98US-0095329.		
PR	10-AUG-1998;	98US-0096012.		
PR	11-AUG-1998;	98US-0096143.		
PR	11-AUG-1998;	98US-0096146.		
PR	12-AUG-1998;	98US-0096229.		
PR	17-AUG-1998;	98US-0096757.		
PR	17-AUG-1998;	98US-0096756.		
PR	17-AUG-1998;	98US-0096758.		
PR	17-AUG-1998;	98US-0096773.		
PR	17-AUG-1998;	98US-0096791.		
PR	17-AUG-1998;	98US-0096867.		
PR	17-AUG-1998;	98US-0096891.		
PR	17-AUG-1998;	98US-0096894.		
PR	17-AUG-1998;	98US-0096895.		
PR	17-AUG-1998;	98US-0096897.		
PR	18-AUG-1998;	98US-0096749.		
PR	18-AUG-1998;	98US-0096850.		
PR	18-AUG-1998;	98US-0096857.		
PR	18-AUG-1998;	98US-0096860.		
PR	18-AUG-1998;	98US-0097022.		
PR	19-AUG-1998;	98US-0097144.		
PR	20-AUG-1998;	98US-0097618.		
PR	24-AUG-1998;	98US-0097661.		
PR	26-AUG-1998;	98US-0097659.		
PR	26-AUG-1998;	98US-0097952.		
PR	26-AUG-1998;	98US-0097954.		
PR	26-AUG-1998;	98US-0097955.		
PR	26-AUG-1998;	98US-0097971.		
PR	26-AUG-1998;	98US-0097774.		
PR	26-AUG-1998;	98US-0097978.		
PR	26-AUG-1998;	98US-0097986.		
PR	26-AUG-1998;	98US-0098014.		
PR	31-AUG-1998;	98US-0098525.		
PR	16-SEP-1998;	98US-0100634.		
PR	12-JAN-1999;	99US-00115565.		
XX	(GENTECH INC.)			
PA	Baker, K., Chen, J., Goddard, A., Gurney, M., Smith, V., Watanabe, CK,	QY	CCAACTTCCTTCTGTGGTGGCAACCTCTGGTGGCTTACAGGACACC	660
PI	Wood WI, Yuan J;	DB	601 CCAAATTCCTTCTGTGGTGGCTTACAGGACACC	660
XX	Claim 2; FIG 150; 822pp; English.	QY	661 AGCCCATGTCAGGTTCCCTGGAGGCCATCTTGGCTTACAGGACACC	780
CC	The invention provides membrane-bound PRO polypeptides and	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	polynucleotides encoding them. The PRO sequences of the invention were	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	identified based on extracellular domain homology screening. The PRO	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	sequences have homology with proteins including LDL receptors, TIE	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	ligands and various enzymes. The membrane-bound proteins and receptor	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	molecules are useful as pharmaceutical and diagnostic agents. Receptor	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	immunoaffinities, for instance, can be used as therapeutic agents to block	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	receptor-ligand interactions. The membrane-bound proteins can also be	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	employed for screening of potential peptide or small molecule inhibitors	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	of the relevant receptor/ligand interaction. The PRO encoding sequences	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	are useful as hybridization probes, in chromosome and gene mapping and	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	the generation of antisense RNA and DNA. PRO nucleic acid sequences	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	will also be useful for the preparation of PRO polypeptides, especially	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	
CC	by recombinant techniques.	781 GCGCTCTGAGTGAAGGAGGTTGGGGTCAAGTAGGAAATGGCCAGGAC	840	

Db	961	TCTGCCAACAAAAAAAGAAAAAA 991	QY	157 CCTACAAGTCCTGGAAGTTTCCCAAAGGCCGTGGTCTCATAACCTGCTGTGCAC 216
RESULT 13			Db	1121 CCTACAAGTCCTGGAAGTTTCCCAAAGGCCGTGGTCTCATAACCTGCTGTGCAC 180
ARH64779		standard; cDNA; 1047 BP.	QY	217 CCCAGCACACCGCCATCACCTATCCCTCTGTGAAACCAGAAATCAAGTGGCA 276
ID	ARH64779		Db	181 CCCAGCACACCGCCATCACCTATCCCTCTGTGAAACCAGAAATCAAGTGGCA 240
XX			QY	277 AGAGGTGTGAGACCAGGCCGCTCCCTCACCTCAAGTCAACTGAAGTCCA 336
AC	ARH64779;		Db	241 AGAGGTGTGAGACCAGGCCGCTCCCTCACCTCAAGTCAACTGAAGTCCA 300
XX			QY	337 GTCCAGACTGCTCACCTACCTCTGCCGCCCCCTCACCTCAAGTCAACTGAAGTCCA 396
DT	11-SEP-2001	(first entry)	Db	301 GTGAGACTGCTCACCTACCTCTGCCGCCCCCTCACCTCAAGTCAACTGAAGTCCA 360
XX			QY	397 GTGCCAGGCTACAGATGCACTGGAGCTGGGAGCTGGTGTGTCAGCTGCGGCCA 456
DE			Db	361 GTGCCAGGCTACAGATGCACTGGGAGCTGGTGTGTCAGCTGCGGCCA 420
XX			QY	457 ACTTCATCTGAGGAGAGGGAGGGAGGCCAGCTGGGAGATGATGATGATGCTGCCTGCCT 516
KW		Human secreted protein cDNA, SEQ ID NO: 55.	Db	421 ACTTCATCTGAGGAGAGGGAGGGAGGCCAGCTGGGAGATGATGATGCTGCCTGCCT 480
OS		Homo sapiens.	QY	517 CGCGAGGCCACCTATACCAACAGCCTGATGGAAAGGATGCCAGTCACCTGAGC 576
OS		Homo sapiens.	Db	481 CGGGAGCCACCTATACCAACAGCCTGATGGAAAGGATGCCAGTCACCTGAGC 540
PN	WO200142451-A2.		QY	577 AGAGACATGCCACGCCAGCTGCCAACCTCTCCCTCCGAGCAGACATGGACT 636
XX			Db	541 AGAGACATGCCACGCCAGCTGCCAACCTCTCCCTCCGAGCAGACATGGACT 600
PD	14-JUN-2001.		QY	637 GTTTCTGTGCCAGGTGCAAAACAGCCATTGCAAGCGCCTCAAGTGTCG 696
XX	07-DEC-2000;	2000W0-IB01938.	Db	601 GTTTCTGTGCCAGGTGCAAAACAGCCATTGCAAGCGCCTCAAGTGTCG 660
PF			QY	697 CC----- 698
XX	08-DEC-1999;	99US-0169629.	Db	661 CCCCAGGAGGGTGGCCAGGCCACCTCTCCCTCCGAGCAGCTCCCTCACTG 720
PR	06-MAR-2000;	2000US-0187470.	QY	699 ----- 719
PR			Db	721 CGGCCATCACCTCCAGATGCTGGGTGGACACCTGGCCAGGTGACAGATG 780
XX			QY	720 GAGGACTGCCAGGGCCTCCCTGGAGGCCCATCTCTGCTGACAGGCC 779
PA	(GEST)	GENSET.	Db	781 GAGGACTGCCAGGGCCTCCCTGGAGGCCCATCTCTGCTGACAGGCC 840
XX	Dumas Milne Edwards J,	Bougueret L, Jobert S;	QY	780 CGCCGTCGTGACTGAGTGANAGAGGTTGGGGTTCAAGATAAGGGAGGTGAGGA 839
PT			Db	841 CGCCGTCGTGACTGAGTGANAGAGGTTGGGGTTCAAGATAAGGGAGGTGAGGA 900
PT		Full length GENSET human nucleic acids encoding potentially secreted proteins, useful in gene therapy and vaccination against a variety of PT diseases, and for diagnosis of those diseases -	QY	840 CGCAAAGCAGGAGCCATGATGAGTAATGAAACCGTCCAGAGCCAAAGGGAGGTGAGGA 899
PT			Db	901 CGCAAAGCAGGAGCCATGATGAGTAATGAAACCGTCCAGAGCCAAAGGGAGGTGAGGA 960
XX			QY	900 AGGCCATCAGGGACTGTTTGGAGTGGAAATGGGAGGTGAGGA 959
DR	2001-367870/38.		Db	961 AGGCCATCAGGGACTGTTTGGAGTGGAAATGGGAGGTGAGGA 1020
DR	PSDB; AAGB9176.		QY	960 CTCTTGCAAAAAAAAAAAAAAAA 986
XX			Db	1021 CTCTTGCAAAAAAAAAAAAAAAA 1047
CC		Claim 7; Page 620-621; 92:pp; English.	RESULT 14	
CC		The invention relates to full length GENSET human nucleic acids encoding potentially secreted proteins. The nucleic acids and the polypeptides they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate GENSET gene expression. For example, they be used to treat disorders associated with decreased GENSET gene expression by rectifying mutations or deletions in a patient's genome that affect the activity of GENSET or by supplementing the patient's own production of GENSET polypeptides. Conversely, antisense nucleic acid molecules may be administered to down regulate GENSET expression by binding with the cells' own genes and preventing their expression. The sense and antisense nucleic acids may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and hence to determine which patients may be in need of restorative therapy.	AAH99131	
CC		The GENSET polypeptides may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of GENSET polypeptide expression and activity. The present sequence is a GENSET nucleic acid of the invention.	ID	AAH99131 standard; cDNA; 935 BP.
CC		Sequence 1047 BP; 235 A; 324 C; 302 G; 186 T; 0 other;	AC	AAH99131;
SQ		Query Match 85.0%; Score 843; DB 22; Length 1047;	XX	
		Best Local Similarity 90.7%; Mismatches 0; Indels 97; Gaps 1;	DT	12-OCT-2001 (first entry)
Qy	37 AGCAGAGGCCCTAACCCACCCAGGATGGGGCTCCCTGGCTGTTCTGCTGGCGTC	96	Db	1021 CTCTTGCAAAAAAAAAAAAAAAA 1047
Db	1 AGCAGAGGCCCTAACCCACCCAGGATGGGGCTCCCTGGCTGTTCTGCTGGCGTC	60	RESULT 14	
Qy	97 TGGCTGCCAGGAGCTCTCCAAAGGCCAGGGAGAAATTACCCCTGTGGTCTCCATTG	156	AAH99131	
Db	61 TGGCTGCCAGGAGCTCTCCAAAGGCCAGGGAGAAATTACCCCTGTGGTCTCCATTG	120	ID	AAH99131 standard; cDNA; 935 BP.

PA	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen GA, Ruben SM, Komatsoulis G;		
PI			
XX			
DR	WPI: 2000-579483/54.		
XX	P-PSDB; AAB39216.		
DR			
XX	Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -		
PT	Claim 1; Page 363-364; 434pp; English.		
XX			
CC	The polynucleotide sequences given in AAC74223-C74279 encode the human secreted proteins represented in AAB39179-B39226. Sequences encoded by the genes, and also protein sequences with which they share homology. The proteins have activities based on the tissues and cells in which they are expressed.		
CC	Examples of activities include: immunosuppressive; antiarthritic;		
CC	anticirrhematic; antiprofertative; cytostatic; cardiac; vasotropism;		
CC	cerebroprotective; nootropic; neuroprotective; antibacterial; viricide;		
CC	fungicide; and ophthalmological. The human secreted proteins, polynucleotides, antagonists, and agonists of the invention may be useful in the treatment, prevention, and/or diagnosis of various disease, disorders and conditions such as autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to regenerate tissues, maintain organs before transplantation, in storage capabilities. Sequences AAC74214-C74222 and AAB39178 are used during the isolation and characterisation of the genes of the invention.		
XX	Sequence 1047 BP; 238 A; 323 C; 299 G; 184 T; 3 other;		
SQ	Query Match 78.0%; Score 773.6; DB 21; Length 1047; Best Local Similarity 87.1%; Pred. No. 4.4e-182; Matches 941; Conservative 1; Mismatches 0; Gaps 3;		
QY	10 CAGGAACTGAGGGTTCTCATCGCCGAGCAGA-GGCCCTACCCACGGCATGGGG 68	RESULT 16	
Db	1 CAGGAACTGAGGGTTCTCATCGCCGAGCAGGAGGCCCTACCCACGGCATGGGG 60	ABT16794	
QY	69 CTCCCTGGCTTCTGCCTGCTGCAAGCACTTCTCAAGCAAGGAG 128	ABT16794	
Db	61 CTCCCTGGCTTCTGCCTGCTGCTGCCGCTGCTGCCAGRCGTTCTCAAGGAGGGAG 120	AC	
QY	129 GAGAAATTACCCCTGGTCCATGTGCCAACAGTCCTGAAAGTTTCCCACAAAGG 188	DE	
Db	121 GAGAAATTACCCCTGGTCCATGTGCCAACAGTCCTGAAAGTTTCCCACAAAGG 180	Human secreted protein gene sequence - SEQ ID NO 43.	
QY	189 CGCTGGGTGCTATACTGCTGTCACCCAGCCACCACGCCATACCTATTCCTC 248	XX	
Db	181 CGCTGGGTGCTATACTGCTGTCACCCAGCCACCACGCCATACCTATTCCTC 240	KW	
QY	249 TGTGAAACAAACATCACAGTGGCCAGAGGTTGTGAGACCAAGGCCCCCTCC 308	KW	
Db	241 TGTGAAACAAACATCACAGTGGCCAGAGGTTGTGAGACCAAGGCCCCCTCC 300	KW	
QY	309 TTCAACCTCAAAGTCACATCAAGTCAGTCCAGTCACCTACTTGCCGGCG 368	KW	
Db	301 TTCAACCTCAAAGTCACATCAAGTCAGTCCAGTCACCTACTTGCCGGCG 360	KW	
QY	369 TCCTCCACCTTAGTGGCCCATGTGGACAGTCGCACTGAGTCAGTGGTGG 428	PR	
Db	361 TCCTCCACCTTAGTGGCCCATGTGGACAGTCGCACTGAGTCAGTGGTGG 420	PR	
QY	429 TCCAAGCAGTGTCTGAGCTGOGGGCAACTCACTGPPGGAAGGGCAAGGCC 488	PA	
Db	421 TCCAAGCAGTGTCTGAGCTGOGGGCAACTCACTGPPGGAAGGGCAAGGCC 442	PA	
	-----GACTAGGGCGAGGGCG-----	XX	
		PI	
		XX	

therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant) agonists are used in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung or urogenital; (b) immune disorders such as Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Sequence 1047 BP; 238 A; 323 C; 299 G; 184 T; 3 other;

Query Match	78.0%	Score 773.6;	DB 25;	Length 1047;	
Best Local Similarity	87.1%;	Pred. No. 4.4e-182;			
Matches	941;	Conservative	1;	Mismatches	0;
		Indels	138;	Gaps	3;
SQ					
Qy	10	CAGAAACTTACGGGTTCTACCTGCCGAGCAGA	GGCCCTAACCCACGGCCATGGG	68	RESULT 18
Db	1	CAGAAACTTACGGGTTCTACCTGCCGAGCAGG	GGCCCTAACCCACGGCCATGGG	60	ABZ3682
Qy	69	CTTCCCTGGCTGTCTGCCTTCATGCGGAGCTG	TGGTCCAGCAGTCCAAAGGC	128	ID ABZ36822 standard; cDNA; 502 BP.
Db	61	CTTCCCTGGCTGTCTGCCTTCATGCGGAGCTG	TGGTCCAGCAGTCCAAAGGC	120	XX
Qy	129	GAAGAAATTACCCCTGCTCTCCATTGCTCA	AAAGTGCTGGTCCAAAGGC	188	AC
Db	121	GAAGAAATTACCCCTGCTCTCCATTGCTCA	AAAGTGCTGGTCCAAAGGC	180	ABZ36822;
Qy	189	CGCTGGGGGCTCATATACTTGTGCTGCTG	TGGTCCAGCAGTCCAAAGGC	248	XX
Db	181	CGCTGGGGGCTCATATACTTGTGCTGCTG	TGGTCCAGCAGTCCAAAGGC	240	DT 21-FEB-2003 (first entry)
Qy	249	TGTGAAACCAAGACATAAAGTGGCAAGA	AAAGTGCTGGTAAAGCCTG	308	XX
Db	241	TGTGAAACCAAGACATAAAGTGGCAAGA	AAAGTGCTGGTAAAGCCTG	300	DE Human GENSET coding sequence SEQ ID 695.
Qy	309	TTCAACCTAACGTCAACTCAACTCA	ACTGCTGGTCACTTCTGG	368	XX
Db	301	TTCAACCTAACGTCAACTCAACTCA	ACTGCTGGTCACTTCTGG	360	XX
Qy	369	TCTTCCACCTAACGTGGCATGTGG	ACTGGTCACTGCTGGAAAGCTGG	428	XX
Db	361	TCTTCCACCTAACGTGGCATGTGG	ACTGGTCACTGCTGGAAAGCTGG	420	XX
Qy	429	TCCAAAGCCAGTGTCTGAGTGGGGCA	AACTTCACACTCTGAGTGGGG	488	XX
Db	443	TCCAAAGCCAGTGTCTGAGTGGGGCA	AACTTCACACTCTGAGTGGGG	442	XX
Qy	489	AGGTGGGAGATGATCTGCGAGGCTC	CTGGTCAAGACAGGGCAGGCC	548	XX
Db	563	AGGTGGGAGATGATCTGCGAGGCTC	CTGGTCAAGACAGGGCAGGCC	502	XX
Qy	549	GGAAAGGATGGCAAGGTCACTGCA	AGGAAAGCATGCCACCTGCA	608	XX
Db	503	GGAAAGGATGGCAAGGTCACTGCA	AGGAAAGCATGCCACCTGCA	562	XX
Qy	609	TCTTTCCTGCGGAGCCAGACATGGACT	GGTCTGGTCTGGTCTGGTCTG	668	CC The present invention relates to novel GENSET polynucleotides (ABZ36404-ABZ36911) encoding polypeptides.
Db	563	TCTTTCCTGCGGAGCCAGACATGGACT	GGTCTGGTCTGGTCTGGTCTG	622	CC polynucleotides and polypeptides are useful in screening and diagnostic assays for abnormal GENSET expression and/or biological activity. They are also useful for screening of compounds for treating or preventing GENSET-related disorders, such as heavy metal toxicity, cancer, inflammatory diseases, immune disorders, and the neuromuscular, central nervous system (CNS), cardiovascular or gastrointestinal effects of the toxicity -
Qy	659	GTCAGCAAGCAGCCCTAACAGTGGG	CTGGTCTGGTCTGGTCTGGTCTG	697	CC
Db	623	GTCAGCAAGCAGCCCTAACAGTGGG	CTGGTCTGGTCTGGTCTGGTCTG	682	CC
Qy	698	-----	-----	697	CC
Db	683	GTCAGCAAGCAGCCCTAACAGTGGG	CTGGTCTGGTCTGGTCTGGTCTG	742	CC
Qy	698	-----	-----	749	CC
Qy					Sequence 502 BP; 105 A; 135 C; 92 T; 1 other;
Db					Query Match 50.6%; Score 501.6; DB 25; Length 502;
Qy					Best Local Similarity 99.8%; Pred. No. 1.1e-114; Mismatches 0; Conservative Matches 501; Indels 0; Gaps 0;

QY	40	AGAGGCCCTACACCCACGGAGGCAATGGGCTCCTGGGTCTCTGGCTTGTGGCTGGTGTGG 99	PR	07-JUN-2000; 20000US-0209467.
Db	1	AGAGGCCCTACACCCACGGAGGCAATGGGCTCCTGGGTCTCTGGCTGGTGTGG 60	PR	28-JUN-2000; 20000US-021486.
QY	100	CTGOCAGGAGCTCTCCAGGAAGGGAGGAAATTACCCCTGTTGCTTCATGCTCT 159	PR	30-JUN-2000; 20000US-021535.
Db	61	CTGOCAGGAGCTCTCCAGGAAGGGAGGAAATTACCCCTGTTGCTTCATGCTCT 120	PR	07-JUL-2000; 20000US-021647.
QY	160	ACAAAGTCTGGAGTTTCCCAAAGGGCTGGGGTCTCATAAACTGGTGTGCACCCC 219	PR	07-JUL-2000; 20000US-021680.
Db	121	ACAAAGTCTGGAGTTTCCCAAAGGGCTGGGGTCTCATAAACTGGTGTGCACCCC 180	PR	11-JUL-2000; 20000US-0217487.
QY	220	AGCACCCACGCCATCACTATCCCTCTCTGGAAACAAAGAACATCAAGTGCCAAAGA 279	PR	11-JUL-2000; 20000US-0217496.
Db	181	AGCACCCACGCCATCACTATCCCTCTGGAAACAAAGAACATCAAGTGCCAAAGA 240	PR	14-AUG-2000; 20000US-02218390.
QY	280	AGGTGGTGAAGCCCCAGGGCGGCTCCTCAACCTCAAGCTCAACTAGTCAGTC 339	PR	26-JUL-2000; 20000US-0220963.
Db	241	AGGTGGTGAAGCCCCAGGGCGGCTCCTCAACCTCAAGCTCAACTAGTCAGTC 300	PR	26-JUL-2000; 20000US-0220964.
QY	340	CAGACCTGCTCACCTACTTCCTCCGGCTCCACCTAGGTGCCCATTTGGACAGTG 399	PR	14-AUG-2000; 20000US-0224518.
Db	301	CAGACCTGCTCACCTACTTCCTCCGGCTCCACCTAGGTGCCCATTTGGACAGTG 360	PR	14-AUG-2000; 20000US-0224519.
QY	400	CCAGGCTACATGCACTGGAGCTGTGTGCTCAAGGCACTGTCTAGTCGGGGCAACT 459	PR	14-AUG-2000; 20000US-022513.
Db	361	CCAGGCTACATGCACTGGAGCTGTGTGCTCAAGGCACTGTCTAGTCGGGGCAACT 420	PR	14-AUG-2000; 20000US-022514.
QY	460	TCACTCTGAGGAGACAGGGGCCAGGGTGGAGATGATCTGGCAGGGTCTCTCGG 519	PR	14-AUG-2000; 20000US-0225157.
Db	421	TCACTCTGAGGAGACAGGGGCCAGGGTGGAGATGATCTGGCAGGGTCTCTCGG 480	PR	14-AUG-2000; 20000US-0225158.
QY	520	GCAGGCCCACTTACACAAAG 541	PR	14-AUG-2000; 20000US-0225159.
Db	481	GCAGGCCCACTTACACAAAG 502	PR	18-AUG-2000; 20000US-0226168.
<hr/>				
RESULT 19				
ID	AAS34076	standard; cDNA; 724 BP.	PR	23-AUG-2000; 20000US-0227182.
XX	AAS34076;		PR	23-AUG-2000; 20000US-0227182.
AC			PR	30-AUG-2000; 20000US-0228924.
XX			PR	01-SEP-2000; 20000US-0228287.
DT	17-DEC-2001	(first entry)	PR	01-SEP-2000; 20000US-0229343.
XX			PR	01-SEP-2000; 20000US-0229344.
DE		Human cDNA encoding a novel foetal antigen, SEQ ID No 600.	PR	01-SEP-2000; 20000US-0229345.
XX			PR	05-SEP-2000; 20000US-0229309.
XX			PR	08-SEP-2000; 20000US-0229513.
XX			PR	06-SEP-2000; 20000US-0230437.
XX			PR	06-SEP-2000; 20000US-0230438.
XX			PR	08-SEP-2000; 20000US-0231242.
XX			PR	08-SEP-2000; 20000US-0231243.
XX			PR	08-SEP-2000; 20000US-0231244.
XX			PR	08-SEP-2000; 20000US-0231413.
XX			PR	08-SEP-2000; 20000US-0231414.
XX			PR	14-SEP-2000; 20000US-0232080.
XX			PR	08-SEP-2000; 20000US-0232081.
XX			PR	12-SEP-2000; 20000US-0231968.
XX			PR	14-SEP-2000; 20000US-0232239.
XX			PR	14-SEP-2000; 20000US-0232398.
XX			PR	14-SEP-2000; 20000US-0232399.
KW		Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;	PR	14-SEP-2000; 20000US-0232401.
KW		immunomodulator; cardiovascular; cytotoxic; nephrotoxic;	PR	14-SEP-2000; 20000US-0233063.
KW		cardiovascular; autoimmune disease; rheumatoid arthritis;	PR	14-SEP-2000; 20000US-0233064.
KW		hyperproliferative disorder; breast neoplasm; cancer;	PR	14-SEP-2000; 20000US-0233065.
KW		cardiovascular disorder; cardiac arrest; cerebrovascular disorder;	PR	21-SEP-2000; 20000US-0234223.
KW		cerebral ischaemia; angiogenesis; nervous system disorder;	PR	21-SEP-2000; 20000US-0234274.
KW		Alzheimer's disease; infection; ocular disorder; corneal infection;	PR	25-SEP-2000; 20000US-0234997.
KW		wound healing; epithelial cell proliferation; food additive.	PR	25-SEP-2000; 20000US-0234998.
OS		Homo sapiens.	PR	25-SEP-2000; 20000US-0235484.
XX			PR	27-SEP-2000; 20000US-0235834.
PN	WO20015512-A2.		PR	29-SEP-2000; 20000US-0235836.
XX			PR	02-OCT-2000; 20000US-0236802.
PD	02-AUG-2001.		PR	02-OCT-2000; 20000US-0237037.
XX			PR	02-OCT-2000; 20000US-0236367.
XX			PR	02-OCT-2000; 20000US-0237039.
XX			PR	13-OCT-2000; 20000US-0237040.
PR	16-MAR-2000;	20000US-0188874.	PR	13-OCT-2000; 20000US-0239935.
PR	17-MAR-2000;	20000US-0190076.	PR	13-OCT-2000; 20000US-0239937.
PR	18-APR-2000;	20000US-0198123.	PR	20-OCT-2000; 20000US-0240950.
PR	19-MAY-2000;	20000US-0205515.	PR	20-OCT-2000; 20000US-0241221.

DT 23-JAN-2002 (first entry)
 XX Human nervous system related polynucleotide SEQ ID NO 12949.
 XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antisickling; antiaemic; antiarthritic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antilulcer; anticonvulsant; antifungal;
 KW antiparasitic; cardiotonic; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 OS Homo sapiens.
 XX WO200159063-A2.
 XX PD 16-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US011334.
 XX PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0190075.
 PR 19-MAY-2000; 2000US-0198123.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216664.
 PR 07-JUL-2000; 2000US-0216890.
 PR 11-JUL-2000; 2000US-0205515.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218230.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0214518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225215.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 22-AUG-2000; 2000US-0225270.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227709.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229348.
 PR 01-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0234081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233163.
 PR 14-SEP-2000; 2000US-0233164.
 PR 21-SEP-2000; 2000US-0233165.
 PR 21-SEP-2000; 2000US-0233423.
 PR 25-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234398.
 PR 26-SEP-2000; 2000US-0235894.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236167.
 PR 29-SEP-2000; 2000US-0236168.
 PR 29-SEP-2000; 2000US-0236169.
 PR 02-OCT-2000; 2000US-0236170.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 20-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0242221.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-024474.
 PR 08-NOV-2000; 2000US-024475.
 PR 08-NOV-2000; 2000US-024476.
 PR 08-NOV-2000; 2000US-024477.
 PR 08-NOV-2000; 2000US-02446478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0246528.
 PR 17-NOV-2000; 2000US-0246532.
 PR 17-NOV-2000; 2000US-0246609.
 PR 17-NOV-2000; 2000US-0246610.
 PR 17-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0246614.
 PR 17-NOV-2000; 2000US-0246615.
 PR 17-NOV-2000; 2000US-0246616.
 PR 17-NOV-2000; 2000US-0246617.
 PR 17-NOV-2000; 2000US-0246618.
 PR 17-NOV-2000; 2000US-0246619.
 PR 17-NOV-2000; 2000US-0246620.
 PR 17-NOV-2000; 2000US-0246621.

Db	30	AAATGAGTGTGTTAGCTGCCTTGGCA	1
RESULT 22			
ID	AAX12009/C		
XX	Human biallelic polymorphic DNA fragment EST376246a.		
XX	Polymorphism; biallelic; human; forensic; paternity testing; disease; detection; phenotypic typing; characteristic; infection; hereditary; autoimmune disease; cancer; inflammation; drug; therapy; medicament; treatment; marker; ss.		
XX	Homo sapiens.		
OS	W09830165-A2.		
PN	14-MAY-1998.		
XX	05-NOV-1997; 97WO-US20313.		
PR	06-NOV-1996; 96US-0030455.		
XX	(WHED) WHITEHEAD INST BIOMEDICAL RES.		
PA	Hudson T, Lander ES, Wang D;		
PT	WPI: 1998-286974/25.		
XX	New isolated nucleic acid segments from the human genome - used for determining polymorphic forms for use in e.g. forensics, paternity testing or phenotypic typing for disease		
PS	Claim 1; Page 212; 310PP; English.		
XX	AAX10269-X12937 are human DNA fragments which contain biallelic polymorphic markers which have been isolated using the primers represented in AAX09121-X10269. The base occupying the polymorphic site is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments can be used in methods for determining polymorphic forms in an individual for use in e.g. forensics, paternity testing or for phenotypic typing for diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, poly cystic kidney disease, hereditary spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, acute intermittent porphyria, autoimmune diseases, inflammation, cancer, diseases of the nervous system, infection by pathogenic microorganisms, and characteristics such as longevity, appearance (e.g. baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments. The isolated polymorphic acid segments can also be used to produce medicaments for the treatment or prophylaxis of such diseases.		
XX	Sequence 150 BP; 30 A; 46 C; 33 G; 40 T; 1 other;		
SQ	Query Match 15.1%; Score 149.6; DB 19; Length 150; Best Local Similarity 99.3%; Pred. No. 1.6e-27; Matches 149; Conservative 1; Mismatches 0; Indels 0; Gaps 0;		
Qy	820 GGAATGGGGAGGACTGGAGGCCATCAGCGTGCACCTTGAGTTTCGTTGAGTTCATGCA	879	
Db	150 GGAAATGGGGAGGCTCAGGGCAAGCAGCCATGAACTGGAGTTCATGAACTGGAGTCATGCA	91	
Qy	880 CAAGCACGGAGGACTGGAGGCCATCAGCGTGCACCTTGAGTTTCGTTGAGTTCATGCA	939	
Db	90 CAAGCACGGAGGACTGGAGGCCATCAGCGTGCACCTGGAGTTTCGTTGAGTTCATGCA	31	

PR 16-AUG-2001; 2001US-312903P.
 PR 10-SEP-2001; 2001US-318962P.
 PR 12-SEP-2001; 2001US-318770P.
 PR 27-SEP-2001; 2001US-325410P.
 PR 27-SEP-2001; 2001US-325681P.
 PR 18-OCT-2001; 2001US-330380P.
 PR 31-OCT-2001; 2001US-335201P.
 PR 14-NOV-2001; 2001US-335201P.
 PR 14-NOV-2001; 2001US-332172P.
 PR 14-NOV-2001; 2001US-332272P.
 PR 21-NOV-2001; 2001US-333184P.
 PR 03-DEC-2001; 2001US-332172P.
 PR 03-DEC-2001; 2001US-332094P.
 PR 04-DEC-2001; 2001US-332092P.
 PR 03-JAN-2002; 2002US-345705P.
 PR 07-MAR-2002; 2002US-0092900.
 XX PA (CURA-) CURAGEN CORP.
 XX PI Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CBA, Li L;
 PI Zerhusen BD, Gusev V, Ji W, Gorman L, Miller CE, Kekuda R;
 PI Paturjan M, Gangolli E, Vernet CAM, Guo X, Tchernev V;
 PI Fernandes ER, Casman SJ, Malanykar UM, Gerlach V, Liu Y;
 PI Anderson D, Spaderna SK, Catterton E, Burgess C, Leite M, Zhong H;
 PI Alsobrook JP, Lepley DM, Rieger DK;
 XX DR WPI; 2002-723332/78.
 XX PT NOXV polypeptides and polynucleotides, useful for preventing or
 PT treating a disorder associated with aberrant NOXV expression or
 PT activity e.g., cancer, hypertension, atherosclerosis, cardionyopathy or
 XX bronchial asthma -
 PS Claim 13; Page 368-369; 1103pp; English.
 XX CC This invention describes novel human NOXV polypeptides which have
 CC cytosatic, cariant, anti arteriosclerotic, antisthmatic and
 CC hypotensive activity. Pharmaceutical compositions comprising the NOXV
 CC proteins or nucleic acid molecules or NOXV antibodies are useful for
 CC preventing or treating a disorder associated with aberrant NOXV
 CC expression or activity e.g., cancer, hypertension, atherosclerosis,
 CC cardionyopathy or bronchial asthma. The products of the invention can
 CC be used for gene therapy or in a vaccine. ABX97008-ABX7185 are cDNA
 CC fragments amplified and isolated by the PCR primers and probes
 CC represented in ABX97186-ABX7593. ABX97008-ABX97185
 XX encode the NOXV proteins described in ABU65041-ABU65218.
 SQ Sequence 1272 BP; 182 A; 356 G; 213 T; 0 other;
 Qry Query Match 4.8%; Score 47.2; DB 24; Length 1272;
 Db Best Local Similarity 48.8%; Fred. No. 0.084;
 Matches 127; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
 Qy 256 CGAAGACATTAAGGGCCAGAAAGTGGTGAAGCCCCAGAGCGGCCCTCCCTCAACC 315
 Db 711 CCCCGAACACAAAGCTGATCGTGTGGAAACCCGTGACCCGCAATCTCGACT 770
 Qy 316 TCAACCTCTCACACTCAAGTCAGTCCAGACAGCTCACCTACTTCTGCGGGCGTCTCCA 375
 Db 771 AGGCGCTGGGCCCGCCGGAGACGCTGGAGCTGGTGGAGCTGGTCAAGC 435
 Qy 376 CCTCAGTGTGCCTGAATGGCAGGTACAGTGACTGGAGCTGGTCAAGC 435
 Db 831 AGGGCTCTGGGCCCGCCGGAGACGCTGGAGCTGGTGGAGCTGGTCAAGC 890
 Qy 436 TAGTGTGTGACCTGGCGCAACTTCACTTGCGAGGACAGAGGGCGAGGGCG 495
 Db 891 ACCTGGACCACTGGCTGCGTACTTCCCCTGTCGACTGCGAGGAGGAGC 950
 Qy 496 AGATGATCTGCCAGGGGTGCC 515

Db 951 GTCAGTCAGCGAACCGGGCC 970

RESULT 24
 ABQ76896 standard; DNA; 53226 BF.

ABQ76896;

XX AC AC

XX DT 13-MAR-2003 (first entry)

XX DE Human G-protein coupled receptor DNA SEQ ID 3.

XX KW G-protein coupled receptor; secretin receptor subfamily; human; SNP;
 GPCR; protease; Parkinson's disease; gene; chromosome X;
 single nucleotide polymorphism; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 3000..50651
 FT /*tag= a
 FT /product= "GPCR"
 FT /note= "this coding sequence is interrupted by
 FT 13 introns"
 FT replace (1746,c)
 FT variation
 FT /*tag= b
 FT replace (1746,g)
 FT variation
 FT /*tag= c
 FT replace (1755,g)
 FT variation
 FT /*tag= d
 FT replace (1761,g)
 FT variation
 FT /*tag= e
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= f
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= g
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= h
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= i
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= j
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= k
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= l
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= m
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= n
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= o
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= p
 FT /note= "SNP, single nucleotide polymorphism"
 FT replace (1761,g)
 FT variation
 FT /*tag= q

RESULT 25
 AAX83003/c standard; DNA; 87350 BP.
 ID AAX83003 standard; DNA; 87350 BP.
 XX OS Homo sapiens.
 XX BN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX FF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE) HYSEQ INC.
 XX FI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR P-PSDB; ABG04149.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
 XX PR Claim 1: SEQ ID NO 4140; 103pp; English.
 XX PS The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, for chromosome polymerase chain reaction (PCR) primers, oligomers, and for sequencing and gene mapping, and in recombinant production of (III). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (III). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide (II) and its binding partners are useful in medical imaging of sites expressing (III). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
 XX Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp://wipo.int/pub/published_pct_sequences.
 XX Sequence 636 BP; 103 A; 246 C; 180 G; 107 T; 0 other;
 XX SQ Query Match 4.7%; Score 47; DB 18; Length 87350;
 XX Best Local Similarity 78.9%; Pred. No. 0.39;
 XX Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 XX 922 GATTTGGAGTCATGCCAAATAATGACTGTGTTAGCTGCTTGCACAAAAA 981
 XX 18894 GATTTGGATTCCTGCTAAATGATGGATTCACTGCTTGCACAAAAA 18835
 XX QY 982 AAAAAAAA 992
 XX Db 18834 ATGGTTAACAA 18824
 XX RESULT 27
 XX ABZB0229 standard; DNA; 249999 BP.
 XX ID ABZB0229
 XX AC ABZB0229;
 XX DT 02-JUN-2003 (first entry)
 XX DE Human tramadol gene region genomic DNA SEQ ID NO:26.
 XX AC Human; neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy;
 XX AC multiple sclerosis; CNS disorder; non-CNS disorder; tramadol
 XX KW nerve injury; neuropathic pain; stroke trauma; non-CNS
 XX DNA encoding novel human diagnostic protein #4140.
 XX AA68336/c standard; cDNA; 636 BP.
 XX AA68336
 XX DT 13-FEB-2002 (first entry)
 XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX DB DNA; chromosomal mapping; gene mapping; gene therapy; forensic;

KW tramordin; human; chromosome 5; gene; ds.
 XX OS Homo sapiens.
 XX WO2003016502-A2.
 XX PD 27-FEB-2003.
 XX PP 21-AUG-2002; 2002WO-US26637.
 XX PR 21-AUG-2001; 2001US-313907P.
 XX PA (MCLA-) MC LAUGHLIN RES INST.
 PI Birmingham JR;
 XX DR 2003-278567/27.
 XX PT New nucleic acid sequence encoding tramordin, e.g., mouse tramid 1, mouse tramid 2, mouse tramid 3, human tramid 1, human tramid 2, human tramid 3 or rat tramid 1, useful for treating CNS, e.g. stroke, multiple sclerosis, trauma, neuropathic pain -
 XX PS Example 6; Fig 9; 177pp; English.
 XX CC The present invention describes an isolated nucleic acid sequence comprising a cDNA sequence encoding mouse tramordin (tramid) 2, mouse tramid 3, human tramid 1, human tramid 3 or rat tramid 1, or the genomic sequence of mouse tramid 1 or mouse tramid 3. Mouse tramid 1 is located to chromosome 11, whereas human tramid 1 is located to chromosome 5q11-33. The tramid sequences have neuroprotective, nootropic, analgesic and cerebroprotective activities, and can be used in gene therapy. The nucleic acid sequences are useful for diagnosing and treating central nervous system (CNS) disorders such as multiple sclerosis, nerve injury, neuropathic pain, stroke or trauma, and non-CNS disorders. The present sequence represents the genomic sequence of the human tramordin gene region, which is given in the exemplification of the present invention.
 XX SQ Sequence 249999 BP; 75050 A; 54012 C; 51931 G; 67805 T; 1201 other;
 XX Query Match Score 44; DB 25; Length 249999;
 XX Best Local Similarity 77.9%; Pred. No. 3.1;
 XX Matches 53; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 XX QY 922 GTATGGGTTTCATCAAAATGAGTGTCTTGCCATTAGCTTAAATGAAACCA 981
 XX DB 75447 GTATTAATAATCCTGTGAATGAGTAATTTAGCTGCTTGGCACAAAAAACCC 75506
 XX QY 982 AAAAAAA 989
 XX DB 75507 CCAAATAA 75514
 XX
 XX RESULT 28
 ABT07673 ID ABT07673 standard; cDNA; 560 BP.
 XX AC ABT07673;
 XX DT 14-NOV-2002 (first entry)
 DE Human breast cancer associated coding sequence SEQ ID NO: 152.
 DE KW Human; breast specific gene; breast specific protein; breast cancer;
 KW gene therapy; cytostatic; gene; ss.
 OS Homo sapiens.
 XX PN WO200264611-A1.
 XX PD 22-AUG-2002.
 XX PR 12-FEB-2002; 2002WO-US4197.
 XX AC ABT07673;
 XX DT 14-NOV-2002 (first entry)
 DE Human breast cancer associated coding sequence SEQ ID NO: 152.
 DE KW Human; breast specific gene; breast specific protein; breast cancer;
 KW gene therapy; cytostatic; gene; ss.
 OS Homo sapiens.
 XX PN WO200264611-A1.
 XX PD 22-AUG-2002.
 XX
 XX FF 12-FEB-2002; 2002WO-US04197.
 XX PR 13-FEB-2001; 2001US-266292P.
 XX PA (DIAD-) DIADEXUS INC.
 PI Salceda S; Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
 PI Sun Y, Liu C;
 XX DR WPI; 2002-657582/70.
 XX PT New breast specific nucleic acids and proteins, useful for identifying,
 PT diagnosing, monitoring, staging, imaging, and treating breast cancer,
 PT and non-cancerous disease states in breast tissue, and in gene therapy
 XX

PS Claim 1; Page 268-269; 367pp; English.

XX The present invention provides human breast specific coding sequences and sequences. These can be used in the diagnosis and treatment of breast cancer and non-cancerous diseases of the breast. The present sequence is a coding sequence of the invention.

XX Sequence 509 BP; 151 A; 109 C; 123 G; 126 T; 0 other;

Query Match Score 4.3%; Score 43; DB 24; Length 509;
Best Local Similarity 77.6%; Pred. No. 0.68;
Matches 52; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 926 TTGGAGTCATGCCAAATGAGTGTTAGCTCTTGCACAAAAA 985
Db 359 TTGGAGTCATACTAAATGAGTAGATTTGCTCTTGCAACATAAACAAAAAT 418

Qy 986 AAAAAA 992
Db 419 GCCACAA 425

RESULT 30
ID AAL38336/C
XX AAL38336 standard; DNA; 143899 BP.
XX AC
DT 15-AUG-2002 (first entry)
XX DE Genomic sequence encoding a human Ngr2 protein.
XX KW Cerebroprotective; cytoprotective; neuroprotective; cytoskeletal; Nogo receptor homologue; NGR2; NGR3; axonal growth; central nervous system; CNS; cerebral injury; spinal cord injury; stroke; demyelinating disease; multiple sclerosis; monophasic demyelinating disease; encephalomyelitis; Marchiafava-Bignami disease; multifocal leukoencephalopathy; panencephalitis; Spomy degeneration; Alexander's disease; Canavan's disease; immune; bait Protein; generic mapping; gene therapy; gene; ds. Krabbe's disease; Canavan's disease; metachromatic leukoencephalopathy; transgenic animal; unregulated cellular growth; cancer; tumour; human; gene; ds.

XX Homo sapiens.
PN WO200229059-A2.
PD 11-APR-2002.
XX 06-OCT-2001; 2001WO-US31488.
PR 06-OCT-2000; 2000US-238361P.
XX (UYA) UNIV YALE.
PA (BIOJ) BIOGEN INC.
XX PI Strittmatter SM, Cate RL, Sah DWY;
XX DR; WPI: 2002-416677/44.
XX Example 2; Page 176-214; 27pp; English.

XX The invention relates to a Nogo receptor homolog polypeptide, NGR2 or NGR3, comprising a 50 amino acid LRRCT sequence, a 284 amino acid NTIRCT sequence, or a 420, 461 or 392 amino acid sequence, all given in the specification. The NGR3 protein or its binding antibody is useful for decreasing inhibition of axonal growth of a central nervous system (CNS) neuron, by contracting the neuron NGR3 or its antibody, and for treating CNS disease, disorder or injury. NGR3 or a vector comprising NGR3 is useful for treating cerebral injury, spinal cord injury, stroke,

CC demyelinating diseases, e.g. multiple sclerosis, monophasic demyelination, encephalomyelitis, multilocular leukoencephalopathy, panencephalitis, Marchiafava-Bignami disease, spongy degeneration, Alexander's disease, Canavan's disease, metachromatic leukoencephalopathy and Krabbe's disease. NGR3 is useful for inducing an immune response in a mammal against NGR3, as a bait protein in a two-hybrid or three-hybrid assay, and as a research tool for identification, characterisation and purification of interacting, regulatory proteins. The nucleotide sequences of the invention are useful for screening for RFLP associated with certain disorders, for genetic mapping, and for gene therapy. The vector containing NGR3 is useful for producing non-human transgenic animals. The NGR3 binding antibody is useful for isolating and purifying NGR3, for localisation and/or quantitation of NGR3, and for diagnostic and therapeutic purposes. The sequences of the invention, vectors and antibodies are useful for treating or preventing unregulated cellular growth such as cancer and tumour growth. This polynucleotide sequence represents the genomic sequence encoding a human NGR2 protein of the invention.

XX Sequence 143899 BP; 36346 A; 35277 C; 35318 G; 35657 T; 1301 other;

SQ Query Match Score 4.3%; Score 42.8; DB 24; Length 143899;
Best Local Similarity 56.1%; Pred. No. 5;
Matches 101; Conservative 0; Mismatches 77; Indels 2; Gaps 1;

Qy 804 GGGGGCTTCAGGATAGGGAAATGGGTAGAGGACGAAAGCAGCAGGCCATGTTAGAAT 863
Db 92710 GGTTGGAGAGAGATGGATGGCTGAAGAACAAAAGGGCAGATGAGCTGATGAGAT 92651

Qy 864 GAAACGGTCCAGAGAGCCAGCACGGCAGGAGCTCGAGCCATCGCCGCACTGTTCTGT 923
Db 92650 GAAACAGCCDPAAGATCTTAATGCAAGCATGCACTAAGTGTAGGTCTATGCTGATGTTG 92591

Qy 924 ATTTGGAGTTGTTATGCAAATTAAGTTGTTAGCTGCTPTGCCAACAAAAAA 983
Db 92590 TTGGGACTT -TCCTAACAGTGTATTTAACTGCTTGTCAARACGGTAGAA 92533

RESULT 31
ID AAF21712/C
XX AAF21712 standard; DNA; 752 BP.
AC AAF21712;
XX DT 27-MAR-2001 (first entry)
XX Human breast and ovarian cancer associated antigen gene SEQ ID 99.
DB Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive; nootropic; neurprotective; antiviral; antiulcer; antiallergic; hepatotrophic; antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibiotic; antifungal; antiparasitic; cardiotonic; immune disorder; Addison's disease; allergy; autoimmune haemolytic anaemia; autoimmune thyroiditis; diabetes mellitus; Crohn's disease; multiple sclerosis; rheumatoid arthritis; ulcerative colitis; cardiovascular disorder; wound healing; neurological disease; ds. Homo sapiens.

XX PN WO200055173-A1.
XX 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI: 2000-611515/58.
DR P-PSDB; AAB58809.

RESULT 33
 AAC77637
 ID AAC77637 standard; cDNA; 2059 BP.
 AC AAC77637;
 XX DT-FEB-2001 (first entry)

XX Human cancer associated gene sequence SEQ ID NO:31.

DB KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW antidiabetic; antihypertensive; vulnery; immunomodulator;
 KW antiinflammatory; antiallergic; antiarthritic; antibacterial; cardiant;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 KW vasotrophic; antipsoriatic; antiangiogenic; Gene therapy; inflammation;
 immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening; ss.
 XX OS Homo sapiens.
 XX PN WO200055350-A1.
 XX PD 21-SEP-2000.
 XX PP 08-MAR-2000; 2000WO-US059882.
 XX PR 12-MAR-1999; 99US-01242470.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Ruben SM;
 XX DR WPI: 2000-587533/55.
 XX P-PSDB; AAB13428.
 XX PT Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer -
 XX PS Claim 1; Page 633-634; 2352pp; English.
 XX CC AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44239. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytosstatic; proliferative; vulnery; immunomodulator; antidiabetic; antihypertensive; antiarthritic; antibacterial; coagulant; dermatological; antiallergic; cardiant; thrombolytic; coagulant; nootropic; vasotrophic; antipsoriatic and antiangiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens AAC78449 to AAC78457 and AAB4240 represent sequences used in the exemplification of the present invention.

XX SQ Sequence 2059 BP; 505 A; 548 C; 552 G; 451 T; 3 other;

Query Match 4.2%; Score 41.8; DB 21; Length 2059;
 Best Local Similarity 61.5%; Pred. No. 2.2;
 Matches 67; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

CC Query 884 CATGGCAAGGACTGAGGGCATAGCTGTGATTGGAACTATGCAAAAT 943

RESULT 34
 ABX63699/C
 ID ABX63699 standard; cDNA; 2672 BP.
 XX AC ABX63699;
 XX DT 26-FEB-2003 (first entry)
 DE Human cDNA #699 differentially expressed in activated vascular tissue.
 KW Human; gene; ss; vascular tissue; cytostatic; atherosclerosis;
 KW carrier; hypotensive; antidiabetic; gynaecological; vasotropic;
 KW cerebroprotective; gene therapy; vascular disease; cancer; coronary;
 KW artery disease; hypertension; diabetes; pre-eclampsia; restenosis;
 KW ischaemia-reperfusion injury; stroke;
 XX OS Homo sapiens.
 XX PN US2002137081-A1.
 XX PD 26-SEP-2002.
 XX PP 08-JAN-2002; 2002US-0044090.
 XX PR 28-JUL-2000; 2000US-222469P.
 PR 08-JAN-2001; 2001US-260483P.
 PA (BAND /) BANDMAN O.
 XX PI Bandman O;
 XX DR WPI; 2003-110597/10.
 XX PT Combination for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, comprises several cDNAs that are differentially expressed in activated vascular tissue -
 XX PS Claim 1; Page 1; 18pp; English.
 XX CC This invention relates to a combination comprising several cDNAs that are differentially expressed in activated vascular tissue. The invention also discloses a high throughput method for detecting differentially expressed cDNAs in a sample. The cDNAs of the invention may have antiarteriosclerotic; cytostatic; cardiant; hypotenetic; antidiabetic; gynaecological; vasotropic; coagulant; antiprotease activities and may be used in gene therapy. The cDNAs of the invention may be used in a high-throughput methods for detecting differential expression of one or more cDNAs in a sample, or screening several molecules or compounds to identify a molecule or compound that specifically binds a cDNA of the invention. A protein encoded by the cDNA may be used to screen several molecules or compounds to identify a ligand that specifically binds to the protein, or to produce or purify an antibody to the protein that can be used to detect a protein in a sample or purify a natural or recombinant protein from a sample. The nucleotides may be useful for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion injury, restenosis, or stroke. The cDNAs can also be used for large-scale genetic or gene expression analysis of several new nucleic acid molecules. Antibodies to the proteins encoded by the cDNAs are useful for diagnosing pre-pathologic disorders, and chronic or acute diseases associated with abnormalities in the sequence of the protein. The present sequence represents a cDNA of the invention that is differentially expressed in

activated vascular tissue.
Note: The sequence data for this patent did not form part of the specification, but was obtained in electronic format directly from USPTO at <http://seqdata.uspto.gov/sequence.html?DOCID=20020137081>.

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH0316 to AAH1282 and
 CC AAH11633 to AAH8742 represent human cDNA sequences; AAB92446 to
 CC AAB9893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

2160	CCGAGAEGGGCCCCGAGAACCGCGCGCCCTGCGCACTGCGCAACCG 2101	SQ	Sequence 3471 BP; 568 A; 1088 C; 1233 G; 582 T; 0 other;
530	TATCACCACAGCTGTATGGAGGTGGTCAACTCGAGAGAGCATGCCA 589	Query Match	4.2%; Score 41.8; DB 22; Length 3471;
21.00	TGCGCAAGAACCGCCTTATGGCATCGGCTTGGGGGAGTAGCGTCTGCACGA 2041	Best Local Similarity	51.3%; Pred. No. 2.6;
590	CAGGCAGGCC 598	Matches	97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
2040	CAGGGAGGC 2032	Matches	97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
INSULT 35			
AAH15559/c	AAH15559 standard; cDNA; 3471 BP.	Qy	410 GATGCACTGGAGCTGTGTTGAGCTGCACTGCACTTGCA 469
		Db	3027 GCTGGCTGGCAATGCTTCAGGGCTGTCTAGCGAGGGAGCTGGCAAGGCC 2968
		Qy	470 GGACAGGGCAGGCCAGGGGGAGATGATGCCAGGGCTCTCGGGAGGCC 529
		Db	2967 CCGAGAEGGGCCCCGAGAACCGCGCGCCCTGCGCACTGCGCAACCG 2908
		Qy	530 TATCACCAAAGCCTGATGGGGAGGATGGCAGTGGCTAACCTGAGAGAGCATGCCA 589
		Db	2907 TGCTACGACACGGCTATTGGCATCGGCTTGGGGGAGAATGCGTCTGCACCGA 2848
		Qy	590 CAGGCAGGCC 598
		Db	2847 CAGGGAGGC 2839
RESULT 36			
		ID	ABL53715 standard; cDNA; 3541 BP.
		XX	ABL53715;
		AC	ABL53715;
		XX	24-JUN-2002 (first entry)
		DI	ABL53715/c
		DE	Human poly(ADP-ribose) polymerase p150 cDNA.
		XX	Poly(ADP-ribose) polymerase; PARP; p150; Parkinson's disease;
		KW	cerebral ischaemia; Alzheimer's disease; Parkinson's disease;
		KW	systemic lupus erythematosus; arthritis; tumour; therapy;
		KW	diagnosis; cytostatic; vasotropic; antidiabetic; nortropin;
		KW	neuroprotective; anti-parkinsonian; antiarthritic; dermatological;
		KW	immunosuppressive; antiinflammatory; expressed sequence tag; EST;
		XX	gene; chromosome 8; ss.
		OS	Homo sapiens.
		XX	WO200222792-A2.
		PH	Location/Qualifiers
		PT	12..3199
		FT	/tag= a
		FT	/product= "Poly(ADP-ribose) polymerase"
		XX	WO200222792-A2.
		XX	
		PD	21-MAR-2002.
		XX	
		PF	11-SEP-2001; 2001WO-EP10494.
		XX	
		12-SEP-2000; 2000EP-0119849.	
Claim 8; SEQ ID 13847; 2537pp + CD ROM; English.			
The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -			
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J; Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;			
WPI: 2001-318749/34.			
Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -			
comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the			

PR 13-JDN-2001; 2001EP-0113577.
 XX (LUESCHER, B.
 PA (AUST) AUSTEN M.
 Luescher B, Austen M;
 XX WPI; 2002-339867/37.
 DR P-PSDB; ABB75711.
 XX New polynucleotide encoding poly(ADP-ribose) polymerase useful as a hybridization probe, and for preparing a pharmaceutical composition for preventing and/or treating tumors -
 PT PT
 PT PS Claim 1; Fig 3; 76pp; English.
 XX The present sequence is that of cDNA encoding a human protein, termed p150, having poly(ADP-ribose) polymerase (PARP) activity. p150 protein was isolated from a native c-Myc complex. Predicted DNA sequences from p150 tryptic peptides were used to screen expressed sequence tag (EST) databases. The present, full-length cDNA was assembled from 2 EST clones. The p150 gene was assigned to chromosome 8 band q24 in close proximity to the c-Myc gene, suggesting a role for the gene in Burkitt's lymphoma, and breast and prostate carcinomas. The p150 protein may participate in activating Myc/Max-driven transcription. It has the capacity to stimulate Myc-mediated transcriptional activation, and the capacity to inhibit cell growth, especially cell proliferation. p150 shows ubiquitous expression, and p150 transcription is elevated in the thymus. The invention provides p150 nucleic acids and polypeptides, expression vectors, host cells, antibodies, pharmaceutical compositions and methods for treating disorders associated with aberrant regulation of cellular behaviour. Use of a p150 antagonist for the treatment of cerebral ischaemia, diabetes, Alzheimer's disease, Parkinson's disease, systemic lupus erythematosus or arthritis, or for the prevention and/or treatment of tumours, is claimed. Also claimed is a diagnostic composition comprising a p150 nucleic acid, protein or antibody.
 XX Sequence 3541 BP; 592 A; 1108 C; 1251 G; 590 T; 0 other;
 SQ Query March Best Local Similarity 4.2%; Score 41.8; DB 24; Length 3541;
 Matches 97%; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
 Qy 410 GATGCACTGGAGCTGTGGTCCAGCCAGTGTGAGTCGGGCCAACTCTACTCTGCA 469
 Db 3079 GCTGGGCTGGCGATGATGGTCCAGGGCTGTTGACAGAGCAGTCGGCAGGACC 3020
 Qy 470 GGACAGAGGGCAGGGCCAGAGATCTGCAGGGCTCCTGGGGCCGCC 529
 Db 3019 CGCAGAGGGGGCGCAAGCCGGCGCAAGCCGGCGCTGCAGTCAGCC 2960
 Qy 530 TATCACAACACCTGATGCGGAAGGATGGAGGTCAACCTGAGAGAACCATGCCA 589
 Db 2959 TGCACAAACACGCCATTATGCCATTGGGGTGGGGGAGTAGGGCTGCCACCGA 2900
 Qy 590 CAGGCAGCC 598
 Db 2899 CAGGGAGGC 2891
 SQ Sequence 38736 BP; 11258 A; 7734 C; 8078 G; 11665 T; 1 other;
 Query Match 4.1%; Score 41; DB 24; Length 38736;
 Best Local Similarity 76.9%; Pred. No. 9-1;
 Matches 50; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 Qy 920 TCGTATTGGAGTTCATGCCAAATGAGTGTTAGCTGCTCTGCAACAAAAAAA 979
 Db 1630 TGTGTTGGCATTTGCTGCAATAGATTTGGTAATGCTGCACTAGATAAAAT 1689
 Qy 980 AAAAA 984
 Db 1690 ACAAA 1694

RESULT 37
 ABQ99952 ID ABQ99952 standard; DNA; 38736 BP.
 XX AC ABQ99952;
 XX DT 12-NOV-2002 (first entry)
 DE Human membrane spanning 4-domain family, subfamily A genomic sequence.
 XX

Human ORFX polynucleotide sequence SEQ ID NO:13317.										
XX	DE	Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis; hyperproliferative disorder; psoriasis; benign tumour; haemorrhage; degenerative disorder; osteoarthritis; neurodegenerative disorder; cardiovascular disease; diabetes mellitus; systemic lupus erythematosus; hypertension; hypothyroidism; cholesterol ester storage disease; immune deficiency; immune disorder; infectious disease; autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis; myasthenia gravis; gene; ss.	XX	KW	KW	KW	KW	KW	KW	KW
XX	ID	ABN97975 standard; DNA; 44100 BP.	XX	AC	ABN97975;	XX	DT	01-AUG-2002 (first entry)	XX	
XX	Human retroviral sequence RAM75.	XX	XX	XX	XX	XX	XX	XX	XX	
XX	Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy; multiple sclerosis; ds.	XX	OS	OS	OS	OS	OS	OS	OS	
XX	Human retrovirus.	XX	XX	XX	XX	XX	XX	XX	XX	
XX	WO9967395-A1.	XX	PN	WO9967395-A1.	XX	PD	29-DEC-1999.	XX	PD	
XX	29-DEC-1999.	XX	XX	23-JUN-1999;	99WO-FR01513.	XX	PF	23-JUN-1999;	PR	
XX	99WO-FR01513.	XX	XX	23-JUN-1998;	98FR-0007920.	XX	PF	23-JUN-1998;	PR	
XX	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.	XX	PA	PA	PA	XX	PI	Shimkets RA, Leach MD;	XX	
XX	Alliel PM, Perin J, Rieger F;	XX	PT	PT	PT	XX	DR	WPI; 2002-106308/14.	DR	
XX	WPI; 2000-160587/14.	XX	PS	PS	PS	XX	PS	Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and autoimmune disorders - Disclosure; SEQ ID 13317: 1037pp; English.	XX	
XX	New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used for diagnosis, treatment and prevention of autoimmune and neurological diseases -	XX	PT	PT	PT	XX	PS	PT Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and autoimmune disorders - Disclosure; SEQ ID 13317: 1037pp; English.	XX	
XX	Claim 15; Fig 14; 225pp; French.	XX	CC	The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification), ABN15762 to ABN22752 encode the human ORFX proteins given in ABP0010 to ABP1150. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmunity inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, repurposing injury in various tissues and conditions resulting from systemic cytokine damage.	CC					
XX	The present invention relates to new nucleic acid sequences of human endogenous retrovirus, HERV-7q, which is located on chromosome 7q. Regulatory elements associated with HERV-7q may alter expression of other genes (even remote genes) on the same chromosome, inducing immunological and/or neurological changes (which may be pathological or protective/curative). HERV-7q peptides can be used to improve efficiency of the immune response, e.g., in immunotherapy. HERV-7q peptides and their coding sequences can be used in immunogenic or vaccination compositions, for protection against autoimmune diseases, particularly multiple sclerosis. The peptides may also be used (by sequence comparison) to detect/identify endogenous retroviruses that are abnormally expressed in cancer, neuropathologies or other autoimmune diseases. The present sequence was used to illustrate the invention.	CC	CC	CC	CC	CC	CC	CC	CC	
XX	Sequence 44100 BP; 13051 A; 8074 C; 8534 G; 14441 T; 0 other;	XX	SQ	Score 4.0%; Best Local Similarity 77.4%; Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	DB	Score 39.4%; Pred. No. 4.4%; Pred. No. 21;	DB	Score 39.4%; Pred. No. 4.4%; Pred. No. 21;	DB	
XX	Sequence 44100 BP; 13051 A; 8074 C; 8534 G; 14441 T; 0 other;	XX	Qy 920 TGTATTTGGGTTCTGAAATGAGTGTGTTTAGCTGCTTCAGCAAAAAAA 979	Db 10427 TTGATTTGGGATTCTATGAGTTAACTACTCTAACCAAAACRAA 10486	Qy 980 AA 981.	DB	Score 40.8%; Pred. No. 4.4%; Pred. No. 21;	DB	Score 40.8%; Pred. No. 4.4%; Pred. No. 21;	
XX	Qy 10487 AA 10488.	XX	Qy 338 TCCGACCTGCTGACCACTACTTGCGGGCTCCACCTCAGGGACAG 397	Db 9 TCCCGCTGCTGAGCCGCTGACCTCTGGCACACACCCCTGGCACCGTGTGATCAT 68	Qy 338 TCCGACCTGCTGACCACTACTTGCGGGCTCCACCTCAGGGACAG 397	DB	Score 50.9%;保守性 0; 匹配数 0; 空缺数 91; 增删数 0; 插入数 0; 缺失数 0;	DB	Score 50.9%;保守性 0; 匹配数 0; 空缺数 91; 增删数 0; 插入数 0; 缺失数 0;	
XX	Sequence 276 BP; 43 A; 105 C; 86 G; 40 T; 2 other;	XX	Qy 338 TCCGACCTGCTGACCACTACTTGCGGGCTCCACCTCAGGGACAG 397	Qy 338 TCCGACCTGCTGACCACTACTTGCGGGCTCCACCTCAGGGACAG 397	Qy 338 TCCGACCTGCTGACCACTACTTGCGGGCTCCACCTCAGGGACAG 397	DB	Best Local Similarity 50.9%; Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	DB	Best Local Similarity 50.9%; Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
XX	Qy 9 TCCCGCTGCTGAGCCGCTGACCTCTGGCACACACCCCTGGCACCGTGTGATCAT 68	XX	Qy 9 TCCCGCTGCTGAGCCGCTGACCTCTGGCACACACCCCTGGCACCGTGTGATCAT 68	Qy 9 TCCCGCTGCTGAGCCGCTGACCTCTGGCACACACCCCTGGCACCGTGTGATCAT 68	Qy 9 TCCCGCTGCTGAGCCGCTGACCTCTGGCACACACCCCTGGCACCGTGTGATCAT 68	DB	Score 40.8%; Pred. No. 4.4%; Pred. No. 21;	DB	Score 40.8%; Pred. No. 4.4%; Pred. No. 21;	

RESULT 40
AAX80672 standard; cDNA; 3285 BP.
 XX
 AAX80672;
 AC
 XX
 DT 22-OCT-1999 (first entry)
 XX
 DB Clone m103_4 encoding secreted protein-m103_4.
 XX
 Secreted protein; cytokine; cell proliferation; immune stimulation;
 KW vaccine; immune suppression; haemopoiesis; tissue growth; activin;
 KW inhibin; chemotaxis; chemokinesis; haemostasis; thrombolytic; ss;
 KW receptor; ligand; anti-inflammatory; cathepsin; tumour; gene therapy.
 XX
 Homo sapiens.
 OS
 FH Key Location/Qualifiers
 CDS 2294..2848
 FT /*tag= a
 FT /product= "Secreted protein m103_4"
 XX
 PN WO9928335-A1.
 XX
 PR 30-NOV-1998; 98US-0203106.
 PR 04-DEC-1997; 97US-0067454.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PP 02-DEC-1998; 98WO-US255512.
 XX
 PR 10-JUN-1999.
 XX
 DR WPI:1999-385352/32.
 DR P-PSDB; AA126037.

The present sequence is a known clone m103_4 (deposited as ATCC 98600) isolated from human foetal kidney cDNA library. It encodes secreted protein m103_4. Recombinant secreted proteins can be produced by transforming host cells and culturing them under suitable conditions. The polynucleotide and protein are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Some predicted biological activities include cytokine and cell proliferation/ differentiation activity, immune stimulating (e.g. a vaccine) or suppressing activity, haemopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemokinetic activity, hemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cathepsin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotide can be used for gene therapy.

XX
 PT New Polynucleotides encoding secreted human proteins
 XX
 PI Agostino MJ, Clark HF, Collins-Racie LA, Evans C;
 PI Fehtel K, Jacobs K, Lavallie ER, McCoy JM, Marberg D;
 PI Steininger RJ, Treacy M, Wong GG;
 XX
 DR WPI: 1999-385352/32.
 DR P-PSDB; AA126037.

Claim 28: Pages 111-112; 124pp; English.

XX
 PT The present sequence is a known clone m103_4 (deposited as ATCC 98600) isolated from human foetal kidney cDNA library. It encodes secreted protein m103_4. Recombinant secreted proteins can be produced by transforming host cells and culturing them under suitable conditions. The polynucleotide and protein are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Some predicted biological activities include cytokine and cell proliferation/ differentiation activity, immune stimulating (e.g. a vaccine) or suppressing activity, haemopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemokinetic activity, hemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cathepsin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotide can be used for gene therapy.

XX
 PS Sequence 3285 BP; 756 A; 813 C; 705 G; 1011 T; 0 other;

Query Match 4.0%; Score 39.4; DB 20; Length 3285;
 Best Local Similarity 72.5%; Pred. No. 10;
 Matches 66; Conservative 0; Mismatches 21; Indels 4; Gaps 1;
 Qy 891 GAGGACTGGGCCATCACGTGCACTTGGAGTCATGCCAAATGAGTG 950
 Db 3049 GAGGAATGTAGTAATAATTGGACT---TGGAGTCCTGTAATGACTAGA 3104
 Qy 951 TTITAGCTCTCTGCAACAAAAAAA 981
 Db 3105 CTTCAGTCCTGCTGCACAAATCCTAAA 3135

RESULT 41
AAS59222 standard; cDNA; 3285 BP.
 XX
 AAS59222;
 AC
 XX
 DT 16-JAN-2002 (first entry)
 XX
 Human cDNA encoding a secreted protein m103_4.
 DE Human
 XX
 KW secreted protein; ss; antiinflammatory; immunosuppressive;
 KW nootropic; neuroprotective; antiarthritic; antimicrobial; vulnerary;
 KW cytostatic; antidiabetic; viricide; antiinfectivity; anticonvulsant;
 KW vasotrophic; antiparkinsonian; immunostimulant; dermatological;
 KW antirheumatic; antitumor; antiulcer; osteopathic; tranquiliser;
 KW cerebroprotective; cytokine; cell prolifera; cell differentiation;
 KW immune deficiency; severe combined immunodeficiency; SCID; tumour;
 KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;
 KW graft-versus-host disease; myelosis; wound healing; ulcer;
 KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; infection; cardiac disease;
 KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;
 KW food supplement.
 XX
 Homo sapiens.
 OS
 PN WO200175068-A2.
 XX
 PR 11-OCT-2001.
 XX
 PD (GEMY) GENETICS INST INC.
 XX
 PP 22-MAR-2001; 2001WO-US09369.
 XX
 PR 30-MAR-2000; 2000US-0539330.
 PR 04-DEC-2000; 2000US-0729674.
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; AAU39004.

Jacobs K, McCoy JM, Lavallie E, Collins-Racie LA, Evans C;
 Treacy M, Agostino MJ, Steininger RJ, Spannling V, Wong GG;
 Clark H, Fehtel K, Marberg D;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; AAU39004.

The invention relates to novel human secreted proteins, the nucleic acids encoding them. The protein may exhibit cytokine, cell proliferation or cell differentiation activity or may induce production of other cytokines in certain cell populations and may exhibit immune stimulating or immune suppressing activity, which is useful for the treatment of various immune deficiencies and disorders e.g. severe combined immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation. The proteins are also useful in the treatment of diseases

XX 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 788.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer; breast cancer; lung cancer; cancer detection; ss. Homo sapiens.

OS WO200102568-A2.

XX PD 11-JAN-2001.

XX PF 30-JUN-2000; 20000MC-US18374.

XX PR 02-JUL-1999; 99US5-0112310.

XX PR 02-JUL-1999; 99US5-0142311.

XX PA (CHIR) CHIRON CORP.

PA (HYSEQ INC.

XX PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhart C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crnkjaek R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LM, Strache-Crain B;

DR WPI; 2001-091805/10.

XX PT Library of polynucleotides for diagnosing a cancerous state of a mammalian cell and detecting cancer, particularly of the colon or prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 657-658; 104pp; English.

XX The present sequence is one of 3351 sequences in a library of human polynucleotides. The library is used to detect differentially expressed genes correlated with a cancerous state of a mammalian cell and can detect colon, prostate, breast and lung cancer. The library can be used to produce probes for detection of mRNA and to produce additional copies of the polynucleotides. The probes can be used for chromosome mapping of the polynucleotide and for detection of transcription levels. Ribozymes or antisense oligonucleotides can be generated. The polynucleotides and their gene products are used as genetic or biochemical markers (e.g. in carcinogenesis pathway and/or monitor the efficacy of therapies and preventive interventions). The polynucleotides, polypeptides and antibodies against them can be used in pharmaceutical compositions to treat the cancers and proliferative disorders such as neoplasia, dysplasia and hyperplasia.

XX Sequence 394 BP; 103 A; 90 C; 115 G; 86 T; 0 other;

Query Match	4.0%	Score	39.2;	DB	22;	Length	394;
Best Local Similarity	66.7%	Pred. No.	5.5;	Mismatches	0;	Gaps	0;
Matches	56;	Conservative	0;				

Qy 909 GCGTGCACTGTCATTGGAGTTCATCAAATGAGTGTITTAAGCTGCTTCGCCA 968

Db 393 GCGGGCGCCCTTGTTGAGACATACACACAGAGTGAATTTAGCTC 334

Qy 969 CAAAAAA_n 992

Db 333 CACAAAAAA_nAGAA 310

XX Sequence 1639 BP; 410 A; 398 C; 357 G; 469 T; 5 other;

Query Match	3.9%	Score	38.8;	DB	21;	Length	1639;
Best Local Similarity	64.4%	Pred. No.	11;	Mismatches	0;	Gaps	0;
Matches	58;	Conservative	0;				

Qy 903 CCATCAGGGACTGTCATTGGAGTTCATGCCAATGAGTGTITTAAGCTCCTC 962

Db 1551 CCAGCACATTATGAACTCAAGTTATATTGCATAAAAGTGGTTATGCCGCTT 1560

Qy 963 TTGCCACAAAAAA_n 992

Db 1561 TTCTCAAAAAAA_nAAA 1590

RESULT 44

AAC98202

ID AA98202 standard; cDNA; 1639 BP.

XX AC AAC98202;

XX DT 09-MAR-2001 (first entry)

RESULT 45

ABK33656

ID ABK33656 standard; cDNA; 2025 BP.

XX AC ABK33656;

DT 08-MAY-2002 (first entry)

with familial hypercholesterolaemia, hypoxanthine-guanine phosphoribosyltransferase associated with Lesch-Nyhan syndrome, phenylalanine hydroxylase (PAH) gene associated with phenylketonuria, dystrophin gene associated with muscular dystrophy, and human cystic fibrosis. The transmembrane conductance regulator gene associated with cystic fibrosis. The present nucleic acid sequence represents a human disease gene sequence that was used in the methods of the invention.

XX Sequence 56737 BP; 15689 A; 11281 C; 11599 G; 18168 T; 0 other;

Query Match Score 38.6; DB 24; Length 5637;
Best Local Similarity 59.9%; Pred. No. 41;
Matches 100; Conservative 0; Mismatches 64; Indels 3; Gaps 2;

QY 819 GCGAATGGGAGCTCAGGGACCCGCAACCGGCCATGGCATGAGTCAGAG 878
Db 22664 GGGAGATGTAGTCAGGATCAAAGTGAATGAGATGACAATCTGAAA 22625

QY 879 CCAAGCACGGCAGGAGCTCAGGCCATCGAGGTGTTGACTGTCAGTGC 938
Db 22624 TATAATGACAAATGAGTATAATGGCT-GTATTGGGGATTCACAC 22566

QY 939 AATATGAGTGTTGTTTAAGCTGCTCTGGCACAAAAAAAGAAAAA 985
Db 22565 TAAATGA--GATTTAACCTCTTGGCACCAAAAGAAAAAA 22521

RESULT 47
ABQ77405 standard; DNA; 185695 BP.
XX ABQ77405;
AC AC
XX DT 10-MAY-2003 (First entry)
XX DE Human THBS1 DNA.
XX Human; THBS1; vascular disease; cardiant; antiarteriosclerotic; stroke; cerebroprotective; gene therapy; coronary artery disease; ischaemia; myocardial infarction; peripheral vascular disease; pulmonary embolism; venous thromboembolism; forensic; paternity testing; GI12583762; gene; SNP; single nucleotide polymorphism; ds.
XX Homo sapiens.
OS XX
FH Key Location/Qualifiers
FT variation replace (5502,t)
FT /*tag= a
FT /standard_name= "SNP"
FT /note= "Single nucleotide polymorphism (ID G334u3)
FT which does not change the THBS1 protein"
XX WO2003016494-A2.
XX PD 27-FEB-2003.
XX 16-AUG-2002; 2002WO-US26343.
XX PR 16-AUG-2001; 2001US-313097P.
PR 05-OCT-2001; 2001US-327485P.
PR 14-DEC-2001; 2001US-0020141.
XX PA (VITI-) VITIVITY INC.
XX McCarthy J, Ableson A;
PI XX
DR WPI; 2003-300617/29.
PR P-PSDB; ABG74673.
XX Identifying a subject as a candidate for a particular course of therapy to treat a vascular disease or disorder, e.g. stroke, myocardial infarction or ischaemia by determining the identity of the nucleotide present at specific positions -

XX PS Claim 3; Fig 11; 568pp; English.

XX This invention describes a novel method for identifying a subject as a candidate for a particular course of therapy to treat a vascular disease or disorder. The method comprises determining the identity of the nucleotide present at specific positions, or their complements, and identifying the subject as a candidate for a particular clinical course of therapy based on the identity of the nucleotide present in that specific position. The method can be used for identifying a subject who is a candidate for further diagnostic evaluation of a vascular disease or disorder and selecting a clinical course of therapy. The products of the invention have cardiant, antiarteriosclerotic and cerebroprotective activity and can be used for gene therapy. The methods disclosed are useful for treating a vascular disease, e.g. atherosclerosis, coronary artery disease, myocardial infarction, ischaemia, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism. The DNA sequences are useful as fingerprint for detecting different individuals within the same species applicable in forensic studies and paternity testing. This sequence encodes the human THBS1 gene represented in GI12583762, used to illustrate the method of the invention.

XX SQ Sequence 185695 BP; 59388 A; 38636 C; 35664 G; 52007 T; 0 other;
Query Match 3.9%; Score 38.4%; Best Local Similarity 53.7%; Matches 101;保守性 0; Mismatches 86; Indels 1; Gaps 1;
DB 112437 GATGTGTGAGACTGGTAGAAATGGGAGATAATAAAAGGATAAAAGTGTGAGAC 112496

QY 795 GAGGAGTTGGGGTTCAGGATAGGGATGGGGTCAGGGAGCAANGCAGGCC 854
DB 112437 GATGTGTGAGACTGGTAGAAATGGGAGATAATAAAAGGATAAAAGTGTGAGAC 112496

QY 855 ATGTAAGATGAAAGTGAAGGCTAGAGCTGAGACTGAGGCTTCAAGTAAATAC 112556
DB 112497 ATGTAAGATGAAAGTGAAGGCTTCAAGTAAATACAGGACTATAGTTAGTAAATAC 112556

QY 915 ACTGTTGCTATTGGAGTTGAGCATGCAAAATGAGTGTCTTGTGCTCTTGGCAAA 974
DB 112557 TGCTTT-TTATTGAGATTCCTGCAAAAGTAGTATGAGTAAATACAGGACTATAGTTAGTAAATAC 112615

QY 975 AAAAAAA 982
DB 112616 CGCAAAAGA 112623

RESULT 48
ABV61090/c
ID ABV61090 standard; cDNA; 294 BP.
XX AC ABV61090;
AC DE Human prostate expression marker cDNA 61081.
XX DT 13-SEP-2002 (First entry)
XX DE Human prostate expression marker cDNA 61081.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacoxyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
PN WO200160860-A2.
XX PD 23-AUG-2001.
PP 20-FEB-2001; 2001WO-US05171.
XX PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189867P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUN-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255288P.
XX

PA	(MILLI-) MILLENIUM PREDICTIVE MEDICINE INC.	PS Disclosure; SEQ ID 1104; 80pp; English.
XX	Schlegel R, Endege WO, Monahan JE;	CC The present invention describes a polynucleotide (I) comprising a CC sequences of a Bifidobacterium genome selected from the nucleotide CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at CC least 90% identity to which hybridises with the sequences given in CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding CC a fusion protein, comprising a sequence selected from 1097 sequences CC given in ABP6258 to ABP6314 ligated in frame to a polynucleotide (I) CC having an amino acid sequence selected from 1097 sequences CC encoding a heterologous polypeptide. (I) has antidiarrhoeal and CC antibacterial activities, and can be used as an inhibitor of Salmonella. CC (II) (which is a probe) is useful for the detection and/or identification CC of Bifidobacterium longum in a biological sample. A carrier containing CC the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) CC can be used for preventing and/or treating diarrhoea brought about by CC pathogenic bacteria and/or rotavirus. The carrier is a food composition CC selected from milk, yogurt, curd, cheese, fermented milks, milk based CC fermented products, ice-creams, fermented cereal based products, milk CC based powders, infant formula, pet food or a pharmaceutical composition CC selected from tablets, liquid bacterial suspensions, dried oral CC supplement, wet oral supplement, dry tube feeding or wet tube feeding. CC (I) is useful in DNA arrays or chips to carry out analysis of the CC expression of the Bifidobacterium genes. ABQ81844 to ABQ81850 represent CC Bifidobacterium related nucleotide sequences given in the Sequence CC Listing from the present invention but not mentioned further within the CC specification. N.B. The sequence data for this patent is not represented in the printed CC specification but is based on sequence information supplied by the CC European Patent Office.
SQ	Sequence 294 BP; 104 A; 26 C; 69 G; 95 T; 0 other;	SQ Sequence 349980 BP; 69195 A; 106952 C; 106128 G; 67705 T; 0 other;
SQ	Query Match 3.8%; Score 38; DB 23; Length 294; Best Local Similarity 81.5%; Pred. No. 9.9; Matches 44; Conservative 0; Mismatches 10; Indels 0; Gaps 0;	SQ Query Match 3.8%; Score 38; DB 24; Length 349980; Best Local Similarity 60.8%; Pred. No. 1e-02; Matches 62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY	939 AAAATGAGTGTGTATTAGCGCTCTTGCCACAAAAA.....AAA 992	QY 255 ACCAGAACATCAAGGTGGCCAAAGAGGTGGTGAAGACGCCACAGAGGTGGCTCTACTCCAGGGGGCTCGTCAAC 314
Db	93 AAAAGGTTTGTATTACCCCCCTTCACCAAA.....AAA 40	Db 222380 ACCAGACCTTAAGGCCACACAAGGGTCAAGCTTACCCAGGGCTCGTCAAC 222439
RESULT 49		
ABQ81848	ID ABQ81848 standard; DNA; 349980 BP.	ABV18588 ID ABV18588 standard; cDNA; 353 BP.
XX	AC ABQ81848;	AC ABV18588;
XX	DT 19-NOV-2002 (first entry)	DT 13-SEP-2002 (first entry)
DE	Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1104.	DE Human prostate expression marker cDNA 18579.
XX	KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial; KW antidiarrhoeic; antibiotic; inhibitor; of Salmonella; defecation; KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria; KW rotavirus; food composition; pharmaceutical composition; gene; ds.	XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; pharmacogenomic marker; gene; BS.
XX	OS Synthetic.	XX OS Homo sapiens.
OS	EP1227152-A1.	XX PN WO200160860-A2.
XX	PR 30-JAN-2001; 2001EP-0102050.	PD 23-AUG-2001.
XX	PR 31-JUL-2002.	XX PF 20-FEB-2001; 2001WO-US05171.
XX	PF 30-JAN-2001; 2001EP-0102050.	XX PR 17-FEB-2000; 2000US-183319P.
XX	(NEST) SOC PROD NESTLE SA.	PR 16-MAR-2000; 2000US-189862P.
XX	DR 2002-668397/72.	PR 25-MAY-2000; 2000US-207454P.
XX	Novel polynucleotide comprising Bifidobacterium genome sequence useful PT as a probe or primer for detecting and/or identifying Bifidobacterium PT longum in a biological sample -	PR 09-JUN-2000; 2000US-211314P.
XX	PT 18-JUL-2000; 2000US-219007P.	PR 13-DEC-2000; 2000US-255281P.
XX	PT	XX

Search completed: February 9, 2004, 13:11:14
Search time: 348 sec